SVENSKA GEOFYSISKA FÖRENINGEN

VOLUME 7, NUMBER 2 **Tellus** MAY 1955

A QUARTERLY JOURNAL OF GEOPHYSICS

Studies on the General Development of Motion in a Two-Dimensional, Ideal Fluid

By PIERRE WELANDER, Institute of Meteorology, University of Stockholm

(Manuscript received March 10, 1955)

Abstract

The purpose of this paper is to discuss qualitatively certain kinds of asymptotic motion in a two-dimensional, ideal fluid by help of methods of statistical mechanics. It is stressed that the final development of such a fluid cannot be adequately described by use of the ordinary equations of motion, but that a "coarse grain" representation should be used. In this representation, the development is characterized by the forming of a single, large vortex accompanied by a certain non-viscous dissipation. The final equilibrium is probably reached almost explosively after a *finite* time. Some experiments which are carried out seem to support this result. In the earlier stages of development we may expect to have some kind of a quasi-equilibrium motion. It is attempted to find conditions under which such a motion can exist, by studies of a point-vortex model.

As known, the hydrodynamic equations of motion cannot be solved analytically to give in closed form the time-development from any given initial state, but we must generally take resort here to some method of numerical integration. Besides obtaining precise numerical results in special cases, we are, however, also interested in getting a more general, qualitative description of the development, comprising features which are characteristic of all cases, or at least the major part of them, irrespective of their initial states.

Essential contribution to such a description is given by the modern theory of turbulence, as developed by Kolmogoroff and others, but the picture is still far from being complete. The intention of this paper is to make a further discussion of some general features with the aid of methods of statistical mechanics. Attempts in that direction have been given earlier by, for instance, BURGERS (1929), TOLLMIEN (1933) and ONSAGER (1949). However, it seems to be a general impression that no decisive advancement can be made along this line until the mathematical apparatus of statistical mechanics has been improved, so that it can effectively tackle continuous systems also. An undertaking of this task is now being made by HOPF (1951) and others, but certainly much remains to be done. Another way would be to abandon wholly the continuum concept and build up the theory from a molecular point of view, and some ideas in that direction will be presented in a later contribution. The general connection between the fields of kinetic theory of gases-hydrodynamics-theory of turbulence is pointed out already in this paper.

The author here wishes to express his thanks to Mr. G. Dahlquist for several stimulating discussions during the course of the investigation.

1. Basic principles of statistical mechanics

Since people working in the field of hydrodynamics are not generally familiar with the

branch of statistical mechanics, it has been found advisable to start with a short discussion of its basic principles.

The theory of classical statistical mechanics is developed for conservative, Hamiltonian systems, which as known are basic in the classical mechanics. The equations of motion for such a system are

$$\left.\begin{aligned} \frac{dp_i}{dt} &= -\frac{\partial W}{\partial q_i} \\ & \qquad (i = 1, 2, \ldots n) \\ \frac{dq_i}{dt} &= \frac{\partial W}{\partial p_i} \end{aligned}\right\} \qquad (1)$$

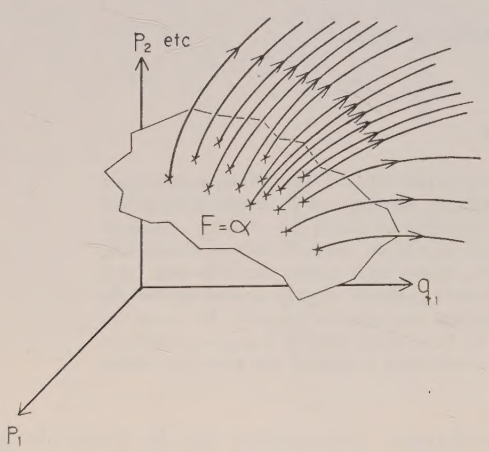


Fig. 1. Experiment "$F = \alpha$".

where $W(p_1, \ldots p_n, q_1, \ldots q_n)$ is the energy of the system, expressed in the coordinates $q_1, \ldots q_n$ and the corresponding "generalized momenta" $p_1, \ldots p_n$. The instantaneous state of a system can be represented as a point, the representative point, in a phase space (or ensemble space) of $2n$ dimensions with the p_i:s and q_i:s as rectangular coordinates. The time development of the system is then represented by a certain trajectory in this space.

Suppose now that we have, in an actual case, only an incomplete knowledge of the initial state of the system. As a specific example, let us assume that we can control at the initial time only a certain function of the initial phase space coordinates, $F(p_1, \ldots p_n, q_1, \ldots q_n)$. If we carry out a number of similar experiments, under steady outer conditions and with a fixed value of F, say α, the initial values p_{i_0} and q_{i_0} are in general found to vary from one experiment to the next due to non-controllable disturbances of our system. However, carrying out a great number of experiments we will find a certain statistical distribution of initial values, and corresponding to this a certain distribution of points on the hypersurface $F(p_1, \ldots p_n, q_1, \ldots q_n) = \alpha$ in the phase space. The whole statistical structure of the experiment "$F = \alpha$" is represented in the phase space by the bundle of trajectories starting at these points (Fig. 1).

Now, it is plausible to assume that the distribution of initial points found on $F = \alpha$ is invariantly connected with our physical system and so is representative also in other experiments with the system, say for instance the experiment "$F = \alpha$ and $G = \beta$". This requires of course that the non-controllable forces ultimately responsible for the statistical spreading of our initial points are not statistically affected by the special arrangement of our experiment. More generally we assume that to every part of the phase space there is related a universal a priori possibility density of initial points of our system. To determine this density distribution is one fundamental task in statistical mechanics.

So far, it has not been necessary to restrict ourselves only to Hamiltonian systems. Such systems, however, have a nice property, which immediately suggests a certain simple distribution of the a priori probability density. Introducing in the phase space the velocity vector $\mathbf{v} = \left(\frac{dp_1}{dt}, \ldots \frac{dp_n}{dt}, \frac{dq_1}{dt}, \ldots \frac{dq_n}{dt}\right)$ we find

$$\nabla \cdot \mathbf{v} = \sum_{i=1}^{n} \left[\frac{\partial}{\partial p_i}\left(\frac{dp_i}{dt}\right) + \frac{\partial}{\partial q_i}\left(\frac{dq_i}{dt}\right)\right] =$$

$$= \sum_{i=1}^{n} \left[\frac{\partial}{\partial p_i}\left(-\frac{\partial W}{\partial q_i}\right) + \frac{\partial}{\partial q_i}\left(\frac{\partial W}{\partial p_i}\right)\right] = 0$$

showing that the flow of representative points in the phase space is an incompressible one. Particularly, this means that a uniform distribution of points throughout the entire phase space is maintained permanently. Considering this result, one finds it natural to assume the existence of a constant a priori probability density in the phase space for all Hamiltonian systems. No rigorous proof of this has been

given, but the hypothesis can be made very plausible also for other reasons, and so far it has never been contradicted by experiments. The hypothesis tells us how to construct our representative "ensemble of systems" in a case where we have only a partial knowledge of the state: the representative points should be distributed so as to all correspond to this partial knowledge, but otherwise they should be distributed uniformly in the phase space.

By use of the theorem of the non-divergent flow and the hypothesis of equal a priori probability density in the phase space one can derive a fundamental theorem, the so-called *H*-theorem. This says that a certain quantity

$$H = \int_{\text{all space}} \dots \int P \log P \, dV$$

where P is the probability density of the representative points and dV stands for the volume element $dp_1\, dp_2 \dots dp_n\, dq_1\, dq_2 \dots dq_n$, will decrease in the run of time towards an absolute minimum value. For the derivation of this theorem reference is given to textbooks (see f.i. TOLMAN: *The Principles of Statistical Mechanics*). For molecular systems, H is proportional to the negative of the entropy, and the *H*-theorem thus expresses the second law of thermodynamics.

The *H*-theorem enables us to determine the state of statistical equilibrium of our system, this being one main task undertaken by the statistical mechanics. The minimum condition of H is in variational form

$$\delta H = \int_{\text{all space}} \dots \int (\log P + 1)\, \delta P \, dV = 0 \quad (2)$$

and to this we should add the subsidiary condition

$$\int_{\text{all space}} \dots \int P \, dV = 1$$

Other subsidiary conditions are given by our partial knowledge of the system. A common case is that we know in the equilibrium state the ensemble mean value of some functions $F_1, F_2, \dots$ depending on the phase space coordinates,

$$\int_{\text{all space}} \dots \int F_i P \, dV = \alpha_i, \quad (i = 1, 2, \dots)$$

Handling these subsidiary conditions by help of Lagrangian multipliers (a_i), we then get for the final probability density the expression

$$P = e^{-(1 + a_0 + a_1 F_1 + a_2 F_2 \dots)} \quad (3)$$

where the constants a_i are determined by reinsertion of (3) into the above conditions. Another frequently occurring case is that some integral to the equations is prescribed, restricting the motion of the systems to a certain hypersurface in the phase space. Most important is here the energy integral, $W(p_1, \dots p_n, q_1, \dots q_n) = \text{constant}$.

Regarding the non-equilibrium states, we have no such general method to fall back on, and it is only in the special branch of kinetic theory of rarefied gases that a complete theory has been worked out. It would be of some interest to discuss here some principal aspects of this theory, which in fact forms the bridge between the statistical mechanics of discrete systems and the hydrodynamic theory of continua.

Starting from the previous phase space representation, the first step in developing the kinetic theory is to introduce the probability distribution function for a single molecule, independent of the states of the other molecules:

$$f(p_1, p_2, p_3, q_1, q_2, q_3) = \\ = \int_{\text{all space}} \dots \int P \, dp_4 \dots dp_n \, dq_4 \dots dq_n$$

where q_1, q_2, q_3 and p_1, p_2, p_3 are the cordinates and momenta, respectively, of the molecule (all molecules regarded as mass-points). For f we can set up a continuity equation, the so-called Boltzmann equation, by equating the net number of molecules disappearing per second in a specified element $dp_1\, dp_2\, dp_3\, dq_1\, dq_2\, dq_3$ to the net number of molecules

a. leaving the q-element due to rectilinear motion.
b. leaving the p-element due to the action of outer forces.
c. leaving the p-element due to molecular collisions.

The contributions according to a. and b. depend directly on f, while the contribution according to c. depends on the joint probability distribution for molecular pairs. To get an equation for f only, we must here

introduce an additional hypothesis, namely that the velocities of the colliding molecules are non-correlated (molecular chaos). If the gas is sufficiently rarefied this hypothesis seems to be plausible enough, however, no rigorous proof has been given.

The second step is to introduce the hydrodynamic variables and stresses as certain mean values computed in the f-distribution: the density ϱ is defined as the mean mass per unit volume, the mass velocity $\mathbf{v}$ as the mean molecular velocity, the normal stress component p_{xx} as the quadratic mean of the difference between the x-components of molecular and mass velocity, etc. Making use of the laws of conservation of mass and momentum in the molecular collisions, we may then derive the hydrodynamic equations of motion. However, these contain as unknowns also the stresses. To get the ordinary Navier-Stokes equation, containing only the hydrodynamic variables ϱ, v and p (where p is defined as the negative mean of the three normal stress components p_{xx}, p_{yy} and p_{zz}), further simplifications have to be made: the molecular velocity distribution must be assumed to deviate only little from the Maxwellian form, and also an additional statistical hypothesis must be introduced (this hypothesis enters the computation very implicitly and cannot be described in simple physical terms).

Examining these two steps we find that they are very similar in nature: in each of the steps the equations are simplified by introducing new dependent variables, defined as certain mean values of the old ones, and at the same time a statistical assumption is being made. From the probability distribution P in the phase space we went over to the single-molecule distribution function f and introduced the statistical hypothesis of non-correlated collisions, and from this we went over to the hydrodynamic variables and introduced again a certain statistical assumption. One can easily understand that the introduction of some hypothesis in each step *is* necessary. As an example, let us assume that we want to compute the further development according to the Boltzmann equation from a given initial field f_0. Now, corresponding to this f_0-field a great number of different distributions P_0 may be constructed. The further developments of these, controlled directly by the Hamiltonian equations, will in general differ, and so will also the corresponding developments of f. To get a *unique* development of f it is then necessary to introduce some restrictive hypothesis, bringing about a direct coupling between the f- and the P fields, and, in fact, such a coupling is yielded by our hypothesis of non-correlated collisions.

Finally, it seems natural to consider the forming of the hydromechanical equations for the mean motion in a turbulent field as the next step in the same sequence. Then a question immediately rises: which statistical hypothesis should be introduced now? This question, which is certainly one of the most fundamental ones in the theory of turbulence, has not yet been answered. However, the recent work by Kolmogoroff and others regarding the quasi-equilibrium state of a turbulent field may eventually provide the basis for a fresh attack on the problem.

2. Development towards a final equilibrium of a two-dimensional, ideal fluid

We consider a two-dimensional, ideal fluid inside a closed and energetically isolated region. At a certain initial time we put the fluid in some arbitrary motion and leave it. How does the motion develop in the long run of time?

First we observe that our system is conservative so that no energy can escape, and thus *some* non-zero asymptotic motion should be found. However, it seems hard to think of a development towards a real equilibrium state. To be able to speak here of an equilibrium approach, we must of course have some irreversible feature in the motion, but our hydrodynamic equations cannot at all distinguish between the two time directions, as is seen by making the transformation $t \to -t$, $\mathbf{v} \to -\mathbf{v}$. In this case every proof, based only on the hydrodynamic equations, that the fluid in the long run approaches some specific equilibrium state can easily be changed to a proof that the fluid instead *departs* from the equilibrium, just by making the above transformation everywhere in the proof.

Now, the same kind of problem has been met earlier in the statistical mechanics, where the basic equations (1) are also time-symmetric. In this case it was found possible to introduce an irreversible feature in the picture, becoming apparent in the famous H-theorem, by

considering not the development of a single system but the development of a whole ensemble of systems starting from various initial states. It should be possible to go the same way in our case, seeing apart from the mathematical difficulties of handling an ensemble of continuous systems. However, we will see that the technique used in the derivation of the *H*-theorem can in fact be utilized here directly for a discussion of the development of the *single system*.

To begin with, let us consider the following simple but illustrative experiment, similar to one discussed by Gibbs in his work *Principles of Statistical Mechanics*. The fluid is initially divided up into a number of surface elements, which are coloured black and white (for instance in a chess-board pattern). The colour should follow the motion materially without any essential diffusion. When setting the fluid into motion the elements are deformed, and after some time they are drawn out in the form of long and thin bands, as shown in Fig. 2.

We now want to investigate the conservation of colour density during the motion. If the black colour is assigned the density 1 and the white colour is assigned the density 0, then the total colour mass of the fluid is $\int_F \varrho\, dF = F_b$, where F is the total area of the fluid, F_b the total area of the black elements. Also the total square density mass is $\int_F \varrho^2 dF = F_b$. Both these quantities are accordingly conservative, since the area F_b is preserved during the motion. However, regarding the total square density mass, the above result is true only if each integration element is either wholly black or wholly white. If we prescribe in advance the size of these elements, however small, we must always ultimately arrive at the situation, where the widths of the bands are found to be so small that the integration elements are each crossed by several bands of different colours. If, as usual, the colour density in an element is defined as the ratio of colour matter and area, then the square density in dF is now $\varrho^{\star 2} = \left[\frac{1}{dF}\int_{dF} \varrho\, \delta F\right]^2 = \left(\frac{dF_b}{dF}\right)^2$, where δF is a "superdifferential" element, and for the total square density mass we have now

$$\int_F \varrho^{\star 2} dF = \int_F \left(\frac{dF_b}{dF}\right)^2 dF = \int_F \left(\frac{dF_b}{dF}\right) dF_b \quad < F_b$$

In this sense, the quantity is not conservative, but decreases. With terms taken from the statistical mechanics, we might call the density $\varrho^\star$ of the fixed integration elements the "coarse grain density", to be distinguished from the "fine grain density" ϱ of the super-differential elements δF, which always are supposed to be small as compared to the widths of the bands.

The argument is now that what we observe actually is not the fine-grain density but the coarse-grain density, the size of the integration elements being always finite due to the existence of an observational limit scale. In the discussed experiment, we have by a direct visual observation a limit scale of some hundredths of a millimeter, and when the widths of the bands have become smaller than this, the "coarse-grain colour" has changed from black or white to grey.

We might expect that our final equilibrium should be characterized by a minimum total square density mass, $\delta \int_F \varrho^{\star 2} dF = 0$, with the subsidiary conditions $\int_F dF =$ constant and $\int_F \varrho^\star dF =$ $=$ constant, giving a uniform (coarse grain) density distribution. The minimum value is here obviously $\frac{F_b}{F} \cdot F_b$. To our ordinary feeling it seems self-evident that a sufficient stirring would ultimately lead to a uniform density distribution, but this feeling may be false. There certainly exist processes which are ordinarily recognized as stirring ones but which do not lead to the uniform density, and it has not yet been proved rigorously that the natural motion of our fluid cannot be such a process. Instead, the ultimate uniformity must be introduced as a hypothesis. However, as far as the two-dimensional and ideal fluid is concerned, we can find support to this hypothesis. The vortex elements of such a fluid form a Hamiltonian system, as will be seen later, and the phase space of this system is identical with the multi-dimensional configuration space. The hypothesis of equal a priori probability density thus says here that each part of the configuration space should be

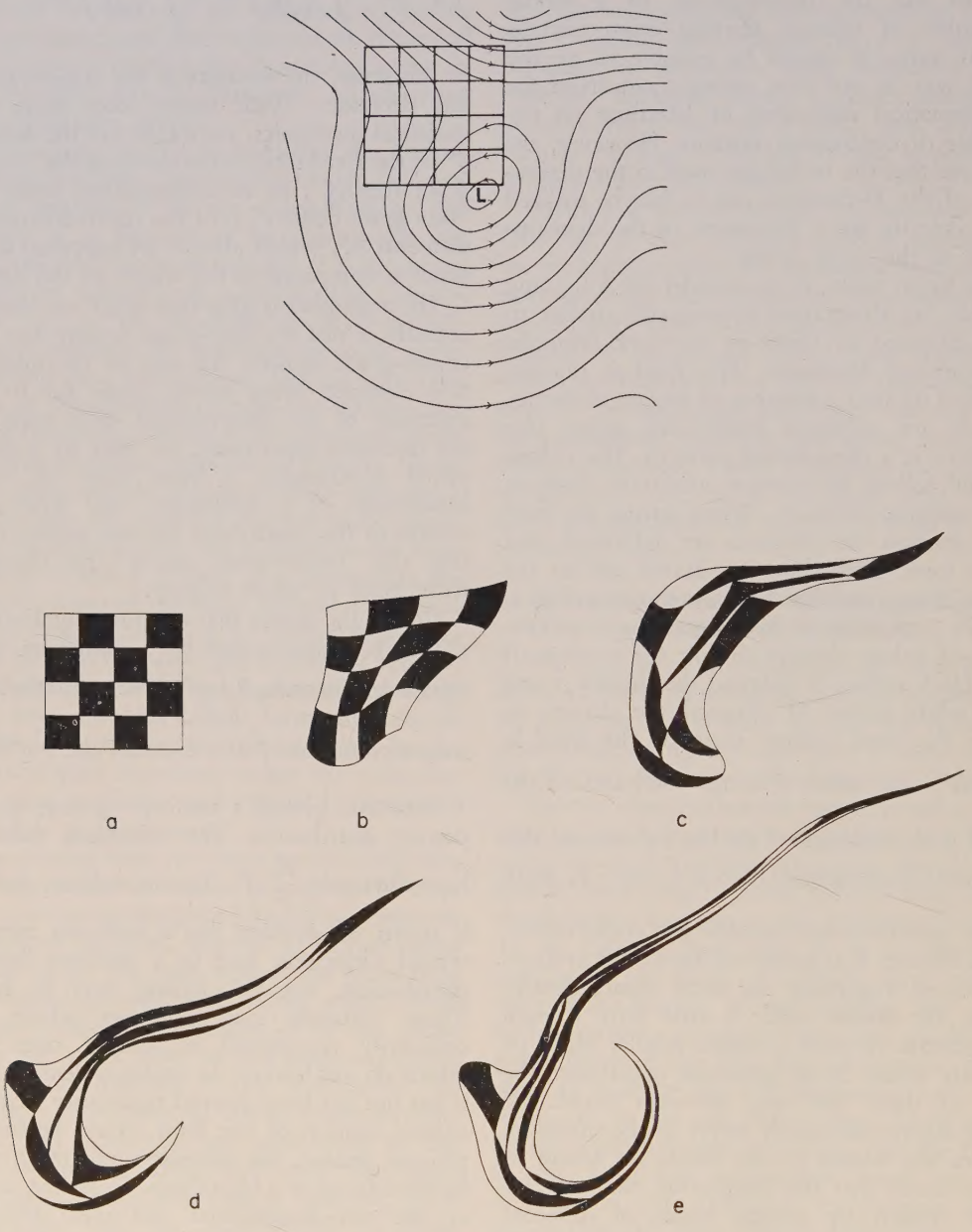


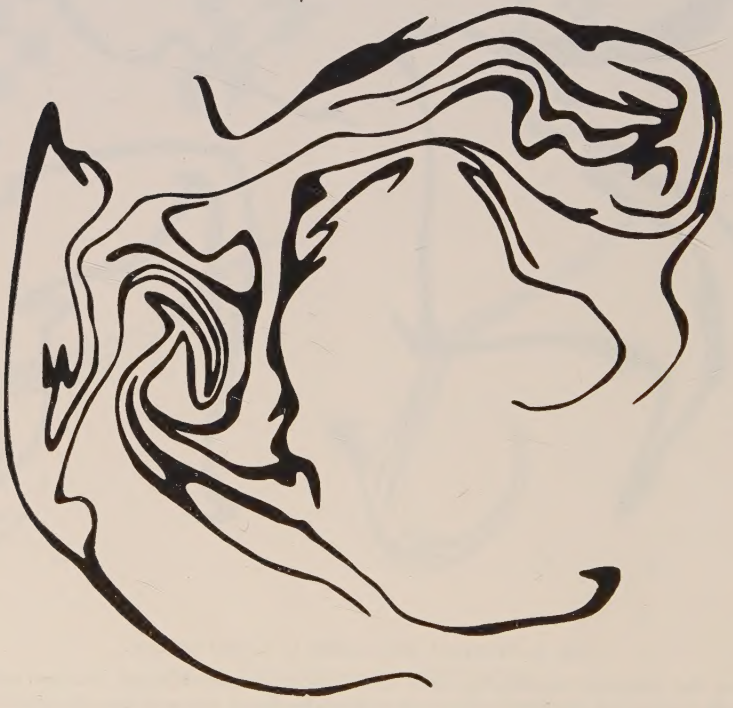
Fig. 2. Deformation of a fluid surface.

The picture shows the deformation of an air layer at the 500 mb level in the atmosphere after 6 h, 12 h, 24 h and 36 h, respectively. The deformation is computed by means of the barotropic model, assuming the layer to behave like a two-dimensional, ideal fluid. The initial stream-line pattern is seen at the top. The sides of the square elements are 300 km. For numerical estimates of the deformation in atmospheric flows see also SAUCIER (1953) and VUORELA (1953).

Fig. 3. Observed deformation of a fluid element.

The picture shows the observed deformation of a small, coloured square element of a fluid surface. A rectangular vessel of dimensions $50 \times 30 \times 30$ cm filled with water to half the depth was used for the experiment. On the water surface was put a film of butanol, which was divided into square elements by means of a metal grid. One or several of these elements were coloured with methyl-red and the water was set into horizontal motion. The grid was then quickly taken away and the fluid was left to move undisturbed. To keep the motion two-dimensional, the whole fluid mass was set into a slow basic rotation before the initial disturbance was created.

f



g

Fig. 3.

given equal statistical weight and this speaks much for the existence of the ultimate uniformity.

Our previous discussion was concentrated on a very special example, but the technique is about the same when considering a more general case. Assuming an arbitrary colour distribution, we just divide the fluid into elements dF so small that we can put initially the density constant within each of these elements. The total coarse grain and fine grain square densities are the same at this time, but at a later time the former quantity is again being diminished, while the latter one is constant. This is seen immediately by inspecting the expressions $\int_F \overline{\varrho^2}^{dF} dF$ and $\int_F (\overline{\varrho}^{dF})^2 dF$, representing these two quantities, respectively (the bar here indicates a mean taken over dF).

From this we turn to the question of the final velocity field. This field is completely determined by the vorticity field of the fluid, and, since the relation between velocity and vorticity is a linear one, the coarse grain velocity field is determined by the coarse grain vorticity field in a corresponding way. Now, the vorticity is in a two-dimensional, ideal fluid a conservative property, which follows the motion of the fluid just in the same way as the colour matter, and the previous conclusions hold as well if we replace the colour density ϱ by the vorticity η. Accordingly, if the stirring is complete, we will ultimately get a state of constant coarse-grain vorticity. The coarse-grain flow is then of the type indicated in Fig. 4, forming a single large vortex.

The total vorticity is conserved here, while a certain amount of square vorticity is lost. This loss is obviously proportional to the mean square fluctuations of the fine grain vorticity in the initial state,

$$-\triangle\left(\int_F \eta^2 dF\right) = \int_F \eta^2 dF - \left[\frac{1}{F}\int_F \eta\, dF\right]^2 \cdot F =$$

$$= \int_F [\eta^2 - (\overline{\eta}^F)^2]\, dF = \int_F (\eta - \overline{\eta}^F)^2 dF =$$

$$= \overline{(\eta - \overline{\eta}^F)^2}^F \cdot F$$

Corresponding to this we also have a certain loss of kinetic energy, disappearing into "turbulent heat".

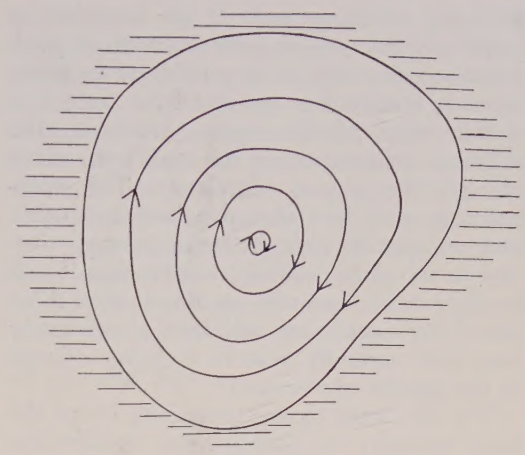


Fig. 4. The final velocity field.

Now, let us stop and try to make the situation clear for us again. In terms of the ordinary fine-grain quantities we cannot describe any final equilibrium state, firstly because of the reversibility which is inherent in the equations of motion, and secondly because of the fact that the velocity field will approach in the run of time a discontinuous state. However, in terms of the coarse-grain quantities an equilibrium is likely to exist, and this equilibrium is here found from the condition of "complete mixing", which is the same as the condition that the mean square vorticity (or the mean square of any other individually conserved quantity) is at the absolute minimum. Two subsidiary conditions hold, namely that the total area and the total vorticity of the fluid is conserved, but these are also the only conservation theorems in the coarse-grain representation. There is one exception, namely when the boundary is exactly circular so that the pressure reactions from the boundary can give no moment around the centre. The angular momentum of the flund is then also conserved with respect to the centre, and since the vorticity of the fluid enters linearly in the angular momentum, this is also conserved in the coarse-grain representation. However, it is to be supposed that the slightest deformation of the boundary will completely change the character of the motion and make it break down again into a complete mixing. We could compare here with the case when a mass-point is moving in a circular region, bounded by a

perfectly reflecting wall. If the boundary is *perfectly* circular, the angular momentum (with respect to the centre) is conserved and the mass-point is restricted to move all the time in a certain zone of the region. However, the slightest deformation of the wall somewhere changes the picture completely. The mass-point is now free to move into any other state compatible with the actual energy value, and in fact it is possible to show that it *will* in the run of time take on almost all of those states[1]. In special, we will after a sufficiently long time certainly observe a change of sign in the angular momentum!

We also want to say something about the time necessary to establish the equilibrium. This time is of course directly dependent on the rate at which the bands are stretched out, and thus we are going to investigate in more detail the mechanism of this stretching. To begin with, it may be good to have a qualitative description of the stretching of a typical band as observed in the experiments (Fig. 3).

Starting from a square element with a side small compared to the scale of the initial motion, we observe first a uniform shearing deformation of the element together with a translation and rotation. Later on the variations of the deformation field over the element become important and the element is deformed more irregularly. After some time a long band is formed, following mainly the streamlines of the fluid, and this is steadily stretched out in the run of time. When the band lies at right angle to the streamlines, folds are found to develop at the points of maximum velocity, also folds appear when the velocity gradient is nearly opposite to the velocity direction. One gets the general impression that the stretching is the natural, stable type of development, while all kinds of shrinkings are unstable, the band immediately bending aside to form a fold which can maintain the stretching tendency. Finally, there have formed a great number of simple and multiple folds and the development then proceeds very quickly to a chaotic state where no detailed features of the band can be observed.

In the theoretical discussion, it is natural to make the assumption that the rate of stretching of a typical band is proportional to the length l of the band itself, the constant of proportionality depending on the features of the velocity field at the actual time,

$$\frac{dl}{dt} \sim \alpha(t)\, l$$

According to the experimental evidence, α should be a positive quantity.

If we have a stationary turbulence field, α should be regarded as a constant and we should find an exponential stretching,

$$l \sim l_0\, e^{\alpha (t - t_0)}$$

This case is considered earlier by BATCHELOR (1952). However, in the case we are discussing, namely the development of the turbulent field from an initial large-scale motion, α will not

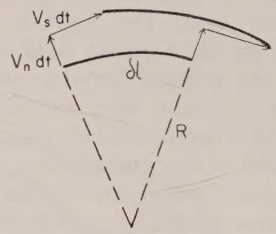


Fig. 5. Stretching of a line element.

be constant but increases all the time, and it seems as if no stationary motion of Batchelor's type exists at all. It seems difficult to make here any precise statement, but we will at least make likely that α increases so rapidly that the bands are stretched out to infinite length and the whole turbulent break-down is completed within a *finite* time.

Consider a small line element of length δl and with a radius of curvature R. During a time dt the length of this element is changed to

$$\delta l + d(\delta l) = \delta l + \frac{\partial v_s}{\partial s}\, \delta l\, dt + \frac{v_n}{R}\, \delta l\, dt$$

where v_s and v_n are the tangential and normal velocity components, respectively (Fig. 5). Thus,

$$\frac{d}{dt}(\delta l) = \left(\frac{\partial v_s}{\partial s} + \frac{v_n}{R}\right) \delta l$$

and integrating over a band from one end-point P_0 to the other P_1, we find for the rate of change of its total length l

[1] The precise statement that can be made here is difficult to express in non-mathematical terms, and the interested readers are referred to text-books on ergodic theory.

$$\frac{dl}{dt} = \int_{P_0}^{P_1} \left(\frac{\partial v_s}{\partial s} + \frac{v_n}{R} \right) ds = v_{s_1} - v_{s_0} + \int_{P_0}^{P_1} \frac{v_n}{R} ds \quad (4)$$

Considering now the experimental results (Fig. 3) we see clearly that the curvature term is large only at certain folds of the band. On the other hand, just at the folds the large curvature term must be approximately balanced by a tangential shrinking,

$$\int_{P_1}^{P_1'} \frac{v_n}{R} ds \sim -(v'_{s_1} - v_{s_1})$$

$$\int_{P_2}^{P_2'} \frac{v_n}{R} ds \sim -(v'_{s_2} - v_{s_2})$$

etc.

(see Fig. 6).

If the curvature term could be neglected except just at the folds, we should then have

$$\frac{dl}{dt} \sim v_0 + 2v_1 + 2v_2 \ldots$$

where v_1, v_2, etc. are the velocity magnitudes at the respective folds. Thus, if V is the characteristic velocity of the fluid, and ν is the actual number of (simple) folds, the order of magnitude of $\frac{dl}{dt}$ is

$$\frac{dl}{dt} \sim 2\nu V$$

Of course, the foregoing estimate is not very satisfactory if the velocity field varies relatively little between the neighbouring folds. Accordingly, folds formed by the small-scale motion of the fluid should not be counted until they have been diffused so far away from each other that they can move independently.

Further support to the assumed mechanism is obtained from a numerical study of the experimental results. Except for a shorter initial period, it is found that about 70—80 % of the stretching is here due to the pure motion of the folds as described above.

Now, $\frac{dv}{dt}$, the number of folds formed per second, should obviously be directly proportional to the total length of the band. Of course, the effective length of the band available for folding is reduced by the existence of earlier folds, but this reduction is precisely balanced by the increased chances for multiple folds. The factor of proportionality should in its turn be inversely proportional to the characteristic length-scale λ of the motion and to a time-constant T, giving the characteristic time required for the developing of a fold, when the conditions for such a development are at hand. T will be something of the order of $\frac{\lambda}{V}$, and thus

$$\frac{dv}{dt} \sim \frac{lV}{\lambda^2}$$

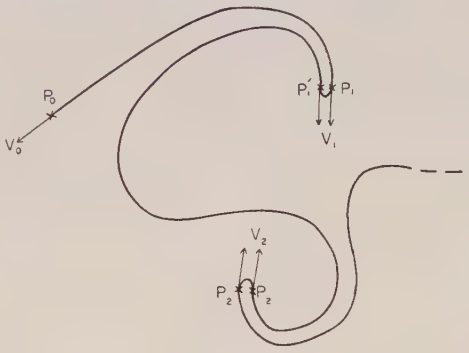


Fig. 6. Stretching of a band.

Here V could be identified with the root-mean square velocity of the fluid, which is conserved during the motion, but λ must be supposed to decrease at the same rate as the width of the band, that is inversely as the length of the band,

$$\lambda \sim \lambda_0 \cdot \frac{l_0}{l}$$

We then find $\frac{dv}{dt}$ proportional to l^3,

$$\frac{dv}{dt} \sim C^2 V l^3$$

where

$$C = \frac{1}{\lambda_0 l_0} .$$

Now,

$$\frac{dv}{dl} = \frac{dv}{dt} : \frac{dl}{dt} \sim \frac{1}{2} C^2 \frac{l^3}{\nu} , \text{ or}$$

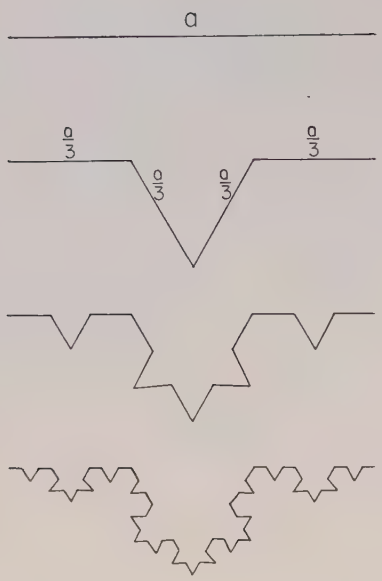


Fig. 7. Construction of the snow-flake curve.

$$\frac{d(v^2)}{dl} \sim C^2 l^3$$

and integrating we have

$$v \sim \frac{C}{2} l^2$$

The number of folds thus should be proportional to l^2.
We get

$$\frac{dl}{dt} \sim CVl^2$$

so that

$$\alpha(t) = CV \cdot l$$

and integrating

$$l \sim \frac{l_0}{1 - C_1(t - t_0)} \qquad (5)$$

where

$$C_1 = Cl_0 V = \frac{V}{\lambda_0}$$

Accordingly, we arrive at the interesting result that the bands are drawn out to infinite length within a *finite* time, which is of the order of magnitude of $\frac{\lambda_0}{V}$, where λ_0 is the characteristic length-scale of the initial motion and V is the characteristic velocity. One might perhaps find the above result unreasonable, since the velocity of the fluid must always remain finite. However, it is certainly possible to deform a finite line-element into a curve of infinite length during a finite time and with finite velocities of all points on the curve. To take a simple example, we consider the steps of displacements shown in Fig. 7, leading to the so-called snowflake-curve.

If the initial length of the line is a, and we maximate all velocities to a value v, then the first step can be made during a time $\triangle t_1 < \frac{1}{3}\frac{a}{v}$, the second during a time $\triangle t_2 < \frac{1}{3^2}\frac{a}{v}$ etc., and the total time to perform any number of steps can be made smaller than $\sum_{k=1}^{\infty} \frac{1}{3^k}\frac{a}{v} = \frac{1}{2}\frac{a}{v}$, while on the other hand the length of the curve after n steps is $a_n = \left(\frac{4}{3}\right)^n a$ and so goes to infinity with n. However, in this example as well as in our fluid case, some space derivatives of the velocity must necessarily go to infinity, making the final velocity field wholly discontinuous.

Of course, the final explosive stretching of the fluid elements can never be observed in reality due to viscous effects, but the above result may be of interest for instance in connection with numerical forecasting, where one tries to predict the development by use of some idealized hydrodynamic model (in fact, indications of a break-down of the motion has been observed in several numerical forecasts with a barotropic model, but it has not yet been made clear whether this is caused by computational errors or not).

It will be stressed here again that our considerations only concern with the *general* (turbulent) development. Of course, we can select *special* cases where the results do not hold, as for instance the cases of stationary motion. The possibility of drawing any general conclusions regarding the development is wholly due to the fact that we exclude certain special and relatively improbable situations.

To make a first crude test of the above theory, the mean stretching of the elements in the described experiment has been measured

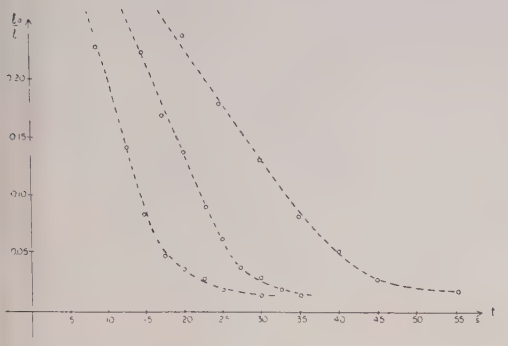


Fig. 8. Observed mean stretching of fluid elements.

(three cases) and plotted inversely against time (Fig. 8). One should not rely too much upon the numerical results, since the three-dimensional and viscous effects disturb the motion in the experiment considerably. However, it is interesting to see how well the points in each case fit a straight line as predicted by (5), at least during the first "inertial period", when the inertia forces still dominate over the viscous forces.

We will end this chapter by saying also a few words about the case of a three-dimensional ideal fluid. The development is here much the same as in the two-dimensional case, but in one respect there is a fundamental difference: the vorticity vectors must no longer be parallel but can turn freely. It has sometimes been argued that since the vorticity vectors are expected to arrive ultimately at a uniform directional distribution, we should find a complete non-viscous dissipation of the motion. However, this needs not to be true, because the magnitude of the vorticity vector and also the anomalies of this will all the time increase due to the stretching of the vortex tubes and the corresponding shrinking of their cross sections. Even if the directional distribution in the run of time becomes uniform, the anomalies may well increase at such a rate that a non-vanishing (coarse grain) velocity field is maintained. Obviously the question cannot be settled before some quantitative analysis has been performed.

3. Quasi-equilibrium motion

The precise development of motion in the foregoing experiment will of course depend on the special velocity field that is set up initially. However, we expect that the details of the initial field should be important only for the motion during a shorter initial period, and that the motion later on should develop into an asymptotic state, the form of which is essentially determined by such over-all quantities as the total energy, the characteristic scale of the motion, etc. Furthermore, it is likely that this asymptotic state could be characterized as a gradually changing quasi-equilibrium. To get a picture of the state, we can imagine the fluid as made up of a number of characteristic, separate fluid elements. The neighbouring elements interact with each other and try to arrive at some mutual equilibrium. On the other hand, they can preserve their individuality only for a shorter time, and are soon broken down into some new elements. However, the time necessary to establish their equilibrium is probably often short as compared to the life-time of the elements and to the time necessary for an essential over-all change of state, and then some kind of a quasi-equilibrium motion is obviously set up.

One can here make a comparison with the kinetic theory of gases. There the development of an initially given molecular distribution has been discussed, and it has been demonstrated that the distribution quickly settles down to a quasi-Maxwellian state, the further changes of which depend only on certain over-all micro-quantities (the hydrodynamic and thermodynamic variables). The time necessary to establish the Maxwellian equilibrium is here not more than some billionths of a second (at normal temperature and pressure), which is much less than the time characterizing an ordinary over-all change of state in the gas.

Concerning the existence of a quasi-equilibrium state in fluid developments, there have been many investigations in the statistical theory of turbulence on the basis of the Kolmogoroff Similarity Hypothesis (see f. i., BATCHELOR 1953), and both from these investigations and from experiments one finds that the quasi-equilibrium concept has a very general applicability. In the statistical theory of turbulence the reasoning centers around the Fourier spectrum and the correlation functions of the energy and velocity fields, and several precise results are obtained, which will not be repeated here. On the other hand it should be

valuable to supplement this essentially statistical line of investigation with a more physical one, discussing by help of the methods of statistical mechanics some simplified mechanical models of the fluid. As a start, we shall here take up to discussion the simple point-vortex model, discussed earlier by Onsager, and try to find out under which conditions this model permits a quasi-equilibrium state to exist.

We consider a number (n) of vortices of strengths $\mu_1, \ldots \mu_n$, moving in an $x-y$-plane. These vortices generate a flow characterized by a stream-function

$$\psi^\star = \sum_{i=1}^{n} \mu_i G(x, y, x_i, y_i)$$

where G is Green's function appropriate to the problem. For an unlimited fluid, which is the only case considered here, we have

$$G(x, y, x_i. y_i) = \frac{1}{2\pi} \log \sqrt{(x-x_i)^2 + (y-y_i)^2}$$

The point-vortices should follow the motion of the fluid materially, thus their velocities are

$$u_k = \frac{dx_k}{dt} = -\left(\frac{\partial \psi^\star}{\partial y}\right)_{\substack{x = x_k \\ y = y_n}} \qquad (k = 1, 2, \ldots n) \quad (6)$$

$$v_k = \frac{dy_k}{dt} = \left(\frac{\partial \psi^\star}{\partial x}\right)_{\substack{x = x_k \\ y = y_k}}$$

We have

$$\left(\frac{\partial \psi^\star}{\partial x}\right)_{\substack{x = x_k \\ y = y_k}} = \frac{\partial}{\partial x_k}\left[\sum_{i=1}^{n} \mu_i G(x_k, y_k, x_i, y_i,)\right]$$

and similarly for $\left(\frac{\partial \psi^\star}{\partial y}\right)_{\substack{x = x_k \\ y = y_k}}$. These last expressions, however, include a (constant) singular contribution representing the infinite self-potential of the vortex-point, and this must be subtracted (rigorously this can be justified by starting from some small vortex-plates and letting their areas tend to zero in the final result). Introducing the energy function

$$W(x_1, y_1, \ldots x_n, y_n) = \sum_{\substack{i, j=1 \\ i>j}}^{n} \mu_i \mu_j G(x_i, y_i, x_j, y_j)$$

we find then that the system (6) can be written in the form

$$\mu_k \frac{dx_k}{dt} = -\frac{\partial W}{\partial y_k}$$

$$\mu_k \frac{dy_k}{dt} = \frac{\partial W}{\partial x_k}$$

Finally, introducing the variables

$$p_k = \sqrt{\mu_k}\, x_k$$

$$q_k = \sqrt{\mu_k}\, y_k$$

this becomes

$$\frac{dp_k}{dt} = -\frac{\partial W}{\partial q_k}$$

$$\frac{dq_k}{dt} = \frac{\partial W}{\partial p_k}$$

The system is thus Hamiltonian and conservative, since W does not depend explicitly on time. In fact ϱW equals the kinetic energy T of the associated fluid, apart from the singular contribution. Accordingly, we have here a case where the methods of statistical mechanics, as described in the first chapter, should apply. There are, however, some special questions to be considered. To begin with, we cannot have a statistically uniform distribution of vortex-points throughout the whole plane, since this would give us infinite velocities. On the other hand, if we consider a local cloud of vortex-points, we will expect that in the run of time the cloud spreads out and at least part of the points escape to infinity. The escape could of course be avoided by introducing a solid boundary around the cloud, but we prefer here to put a restriction on the characteristic scale λ of the system, prescribing its ensemble mean value. The characteristic scale could for instance be the mean distance between two vortex-points picked out at random.

Moreover, some restriction should be put on the energy of the system. In most experiments, we study systems having energies in a fairly narrow range, and an ensemble lacking any restriction in this respect cannot at all be representative. In fact, such an ensemble generally puts an overwhelming weight to the very large energies. Thus we shall prescribe here at least the ensemble mean energy.

Introducing then the three subsidiary conditions

$$\mu_1 \mu_2 \ldots \mu_n \int_{\text{all space}} \ldots \int P \, dV = 1$$

$$\mu_1 \mu_2 \ldots \mu_n \int_{\text{all space}} \ldots \int \lambda P \, dV = \overline{\lambda} \qquad (7)$$

$$\mu_1 \mu_2 \ldots \mu_n \int_{\text{all space}} \ldots \int W P \, dV = \overline{W}$$

where dV stands for a configuration element $dx_1 \, dx_2 \ldots dx_n \, dy_1 \, dy_2 \ldots dy_n$, the minimization of the quantity H (cf. (2)) will lead to the following expression for the equilibrium distribution of probability density

$$P = C e^{-a\lambda} \prod_{\substack{i,j=1 \\ i>j}} r_{ij}^{-\frac{\mu_i \mu_j}{2\pi\Theta}} \qquad (8)$$

C, a and Θ are constants to be determined from the subsidiary conditions (7). C and a are of course both positive, while Θ may be either positive or negative.

We recall here the significance of P: $P\mu_1 \mu_2 \ldots \mu_n \, dx_1 \, dy_1 \ldots dx_n \, dy_n$ is the probability to find a configuration state where the first vortex-point μ_1 lies inside a surface element $dx_1 \, dy_1$ at the point x_1, y_1, and the second vortex-point μ_2 lies inside a surface element $dx_2 \, dy_2$ at the point x_2, y_2, etc.

Now, to get any equlibrium at all we must require that all the previous integrals converge. At infinity the integrals certainly tend toward zero rapidly enough to secure convergence, but some trouble may arise at the point $r_{ij} = 0$. The most critical contribution is of the type

$$\int_{\text{at } r_{ij}=0} \ldots \int \log r_{ij} \, r_{ij}^{-\frac{\mu_i \mu_j}{2\pi\Theta}} dx_i dy_i \, dx_j dy_j$$

Introducing here x_i, y_i as a new fixed origin, we may consider instead

$$\int_{r_{ij}=0}^{\varepsilon} \log r_{ij} \, r_{ij}^{-\frac{\mu_i \mu_j}{2\pi\Theta}} dx_i dy_i \, 2\pi r_{ij} dr_{ij} =$$

$$= 2\pi dx_i dy_i \int_0^{\varepsilon} \log r_{ij} \, r_{ij}^{1-\frac{\mu_i \mu_j}{2\pi\Theta}} dr_{ij}$$

and derive the convergence criterion

$$\mu_i \mu_j < 4\pi\Theta \quad (i, j = 1, 2, \ldots n) \qquad (9)$$

Regarding the sign of Θ, which is of course of great importance, some information could be obtained from a qualitative discussion of the "structure function"

$$\phi(E) = \int_{W>E} \ldots \int e^{-a\lambda} \, dV$$

the integration being taken over that part of the phase space where the energy W is less than a certain value E. The probability to find the system at an energy E, varied over a small range dE, is proportional to

$$e^{-\frac{E}{\Theta}} \frac{d\phi}{dE}(E) \, dE$$

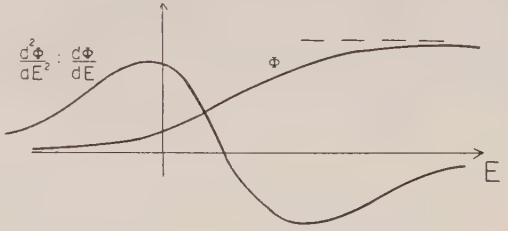


Fig. 9. Appearance of the structure function.

and the maximum probability occurs when

$$\frac{d^2\phi}{dE^2}(E) = \frac{1}{\Theta} \frac{d\phi}{dE}(E)$$

The energy value at which this maximum is attained is expected to lie in the neighbourhood of $\overline{W}$, and $\frac{d^2\phi}{dE^2}(\overline{W}) : \frac{d\phi}{dE}(\overline{W})$ may then be considered as a first approxmation to $\frac{1}{\Theta}$. The general form of ϕ and $\frac{d^2\phi}{dE^2} : \frac{d\phi}{dE}$ is indicated in Fig. 9.

Combining now this with the convergence criterion we can draw the following general conclusions:

1. For a given set of vortices, an equilibrium exists if the energy of the system is sufficiently high or sufficiently low.

2. For a given energy value, an equilibrium exists if the vortices are sufficiently weak.

3. In the non-equilibrium case, the vortices will combine into new vortices, which are intensified if the energy is high but weakened if the energy is low.

In the general case it seems difficult to give some quantitative results, since we cannot evaluate the ϕ-integral and determine the exact form of Θ as depending on E and the μ_i:s. On the other hand, it seems possible to get precise results of interest in some special cases. The discussion of these cases, which involve a lot of mathematical transformations will not be taken up here, however.

Finally we would like to make a remark concerning the physical significance of the model discussed. It is clear that our point-vortex model can represent a given fluid state to any degree of accuracy if the number of point-vortices is increased sufficiently. However, our intention is not to represent here the final equilibrium state of the continuous fluid (in fact, it is not possible to make a transition over the continuous state in our result (8) by making the number of point-vortices infinite!), but to represent the momentary equilibrium attained by some characteristic, finite lumps of fluid. It seems most natural to think of each point-vortex as representing a real, physically separated vortex of the fluid. As an example, in describing the large-scale atmospheric state by our model we should represent each cyclone and anticyclone by one point-vortex of appropriate strength. It cannot be denied that the interpretation is somewhat vague and subjective, but it seems hard to make here any precise statement. On the other hand, considering later on more refined models where the scales of the motion also enter, for instance a model built up of vortex-plates, one should be able to set up more precise connections between the model and the fluid.

REFERENCES

Batchelor, G. K., 1952: The effect of homogeneous turbulence on material lines and surfaces. *Proc. Roy. Soc.* A **213**, 349.

— 1953: *The theory of homogeneous turbulence*, Cambridge Un. Press.

Burgers, J. M., 1923: On the application of statistical mechanics to the theory of turbulent fluid motion. *Vertr. Kon. Akad. v. Wetensch. Amsterdam* **32**, 414.

Gibbs, J. W., 1902: *Elementary principles in statistical mechanics*. In *Coll. Works II*. 1948. Yale Un. Press.

Lamb, H., 1932: *Hydromechanics*, Cambridge Un. Press.

Onsager, L., 1943: Statistical Hydromechanics. *Nuovo Cim., Suppl.* 6, n:o 2, 273.

Saucier, W., 1953: Horizontal deformation in atmospheric motion, *Trans. Am. Geoph. Union*, **34**, 709.

Tolman, R. C., 1938: *The principles of statistical mechanics*. Oxford Un. Press.

Tollmien, W., 1933: Der Burgersche Phasenraum und einige Fragen der Turbulenzstatistik. *Zeits. für Angew. Math. u. Mech.*, **13**, 331.

Vuorela, L., 1953: A synoptic study of deformation fields. *Tellus* **5**, 413.

Available Potential Energy and the Maintenance of the General Circulation

By EDWARD N. LORENZ, Massachusetts Institute of Technology[1], Cambridge, *Mass.*, USA

(Manuscript received Aug. 25 1954)

Abstract

The available potential energy of the atmosphere may be defined as the difference between the total potential energy and the minimum total potential energy which could result from any adiabatic redistribution of mass. It vanishes if the density stratification is horizontal and statically stable everywhere, and is positive otherwise. It is measured approximately by a weighted vertical average of the horizontal variance of temperature. In magnitude it is generally about ten times the total kinetic energy, but less than one per cent of the total potential energy.

Under adiabatic flow the sum of the available potential energy and the kinetic energy is conserved, but large increases in available potential energy are usually accompanied by increases in kinetic energy, and therefore involve nonadiabatic effects.

Available potential energy may be partitioned into zonal and eddy energy by an analysis of variance of the temperature field. The zonal form may be converted into the eddy form by an eddy-transport of sensible heat toward colder latitudes, while each form may be converted into the corresponding form of kinetic energy. The general circulation is characterized by a conversion of zonal available potential energy, which is generated by low-latitude heating and high-latitude cooling, to eddy available potential energy, to eddy kinetic energy, to zonal kinetic energy.

1. The concept of available potential energy

The strengths of the cyclones, anticyclones, and other systems which form the weather pattern are often measured in terms of the kinetic energy which they possess. Intensifying and weakening systems are then regarded as those which are gaining or losing kinetic energy. When such gains or losses occur, the source or sink of kinetic energy is a matter of importance.

Under adiabatic motion, the total energy of the whole atmosphere would remain constant. The only sources or sinks for the kinetic energy of the whole atmosphere would then be potential energy and internal energy.

In general the motion of the atmosphere is not adiabatic. The only nonadiabatic process which directly alters kinetic energy is friction, which ordinarily generates internal energy while it destroys kinetic energy, but which may also, under suitable circumstances, change atmospheric kinetic energy into the kinetic and potential energy of ocean currents and ocean waves. The remaining nonadiabatic processes, including the release of latent energy, alter only the internal energy directly. Hence the only sources for the kinetic energy of the whole atmosphere are atmospheric potential energy and internal energy, while the environment may also act as a sink.

[1] The research resulting in this work has been sponsored by the Geophysics Research Directorate of the Air Force Cambridge Research Center, under Contract No. AF 19 (604)—1000.

It is easily shown (cf. HAURWITZ 1941) that the potential and internal energies within a column extending to the top of the atmosphere bear a constant ratio to each other, to the extent that hydrostatic equilibrium prevails. Hence, net gains of kinetic energy occur in general at the expense of both potential and internal energy, in this same ratio. It is therefore convenient to treat potential and internal energy as if they were a single form of energy. The sum of the potential and internal energy has been called *total potential energy* by Margules (1903).

Evidently the total potential energy is not a good measure of the amount of energy available for conversion into kinetic energy under adiabatic flow. Some simple cases will serve to illustrate this point. Consider first an atmosphere whose density stratification is everywhere horizontal. In this case, although total potential energy is plentiful, none at all is available for conversion into kinetic energy. Next suppose that a horizontally stratified atmosphere becomes heated in a restricted region. This heating adds total potential energy to the system, and also disturbs the stratification, thus creating horizontal pressure forces which may convert total potential energy into kinetic energy. But next suppose that a horizontally stratified atmosphere becomes cooled rather than heated. The cooling removes total potential energy from the system, but it still disturbs the stratification, thus creating horizontal pressure forces which may convert total potential energy into kinetic energy. Evidently removal of energy is sometimes as effective as addition of energy in making more energy available.

We therefore desire a quantity which measures the energy available for conversion into kinetic energy under adiabatic flow. A quantity of this sort was discussed by MARGULES (1903) in his famous paper concerning the energy of storms. Margules considered a closed system possessing a certain distribution of mass. Under adiabatic flow, the mass may become redistributed, with an accompanying change in total potential energy, and an equal and opposite change in kinetic energy. It the stratification becomes horizontal and statically stable, the total potential energy reaches its minimum possible value, and the gain of kinetic energy thus reaches its maximum. This maximum gain of kinetic energy equals the maximum amount of total potential energy available for conversion into kinetic energy under any adiabatic redistribution of mass, and as such may be called the *available potential energy*.[1]

Available potential energy in this sense can be defined only for a fixed mass of atmosphere which becomes redistributed within a fixed region. The storms with which Margules was primarily concerned do not consist of fixed masses within fixed regions, nor do any other systems having the approximate size of storms. It is perhaps for this reason that available potential energy has not become a more familiar quantity.

It is in considering the general circulation that we deal with a fixed mass within a fixed region—the whole atmosphere. It is thus possible to define the available potential energy of the whole atmosphere as the difference between the total potential energy of the whole atmosphere and the total potential energy which would exist if the mass were redistributed under conservation of potential temperature to yield a horizontal stable stratification.

The available potential energy so defined possesses these important properties:

(1) The sum of the available potential energy and the kinetic energy is conserved under adiabatic flow.

(2) The available potential energy is completely determined by the distribution of mass.

(3) The available potential energy is zero if the stratification is horizontal and statically stable.

It seems fairly obvious that the available potential energy so defined is the only quantity possessing these properties, although a rigorous proof would be somewhat involved. Moreover, it possesses the further property:

(4) The available potential energy is positive if the stratification is not both horizontal and statically stable.

It follows from property (1) that available potential energy is the only source for kinetic

[1] This quantity is called *available kinetic energy* by Margules, since it represents an amount of kinetic energy attainable. From the point of view of this discussion the term *available potential energy* is preferable, since it represents a part of the existing total potential energy.

energy. On the other hand, it is not the only sink. When friction destroys kinetic energy it creates internal energy, but in doing so it increases the minimum total potential energy as well as the existing total potential energy. Thus the loss of kinetic energy exceeds the gain of available potential energy.

There is no assurance in any individual case that all the available potential energy will be converted into kinetic energy. For example, if the flow is purely zonal, and the mass and momentum distributions are in dynamically stable equilibrium, no kinetic energy at all can be realized. It might seem desirable to redefine available potential energy, so that, in particular, it will be zero in the above example. But the available potential energy so defined would depend upon both the mass and momentum distributions. If it is desired to define available potential energy as a quantity determined by the mass distribution, the definition already introduced must be retained.

2. Analytic expressions and approximations

Θ	:	potential temperature
p	:	pressure
T	:	temperature
z	:	elevation
$\mathbf{v}$	:	horizontal velocity vector
g	:	acceleration of gravity
c_v, c_p	:	specific heats of air at constant volume and constant pressure
R	:	gas constant for air, equal to $c_p - c_v$
λ	:	ratio of specific heats, c_p/c_v, approximately $^7/_5$
$\varkappa$	:	ratio R/c_p, equal to $\lambda - 1/\lambda$, approximately $^2/_7$

Under adiabatic flow, the statistical distribution of Θ is conserved. More precisely, if $g(\Theta_1)d\Theta$ is the probability that a unit mass of atmosphere selected at random has a value of Θ between Θ_1 and $\Theta_1 + d\Theta$, the probability function $g(\Theta)$ is conserved. If

$$\bar{p}(\Theta_1) = \bar{p}_0 \int_{\Theta_1}^{\infty} g(\Theta)\, d\Theta \qquad (1)$$

where $\bar{p}_0$ is the average value of the pressure p_0 at the earth's surface, regarded as horizontal, $\bar{p}(\Theta)$ is also conserved under adiabatic flow. If the particles for which $\Theta = \Theta_1$ form a continuous surface which intersects every vertical column exactly once, $\bar{p}(\Theta_1)$ is the average pressure on the isentropic surface $\Theta = \Theta_1$, with respect to the area of the horizontal projection of this surface. Equation (1) also defines the average pressure over isentropic surfaces which intersect the ground, or which lie entirely "underground", if along each vertical we define $p(\Theta) = p_0$ if $\Theta < \Theta_0$, where Θ_0 is the value of Θ at the earth's surface.

To express the minimum total potential energy in terms of the invariant $\bar{p}\ (\Theta)$ it is sufficient to express the total potential energy in terms of $p(\Theta)$. The potential and internal energies per unit mass are gz and $c_v T$, respectively. Since, as mentioned previously, the potential and internal energies P and I of a vertical column above a unit area bear the ratio $P/I = (c_p - c_v)/c_v$, and since an element of mass per unit area is $g^{-1}\, dp$,

$$P + I = c_p g^{-1} \int_0^{p_0} T dp \qquad (2)$$

Upon substituting $T = \Theta p^{\varkappa} p_{00}^{-\varkappa}$, where $p_{00} = 1{,}000$ mb, and integrating by parts, we find that

$$P + I = (1 + \varkappa)^{-1} c_p g^{-1} p_{00}^{-\varkappa} \int_0^{\infty} p^{1+\varkappa} d\Theta \qquad (3)$$

The minimum total potential energy which can result from adiabatic rearrangement occurs when $p = \bar{p}$ everywhere, and is obtained by setting $p = \bar{p}$ in (3). Thus the average available potential energy per unit area of the earth's surface is

$$\bar{A} = (1 + \varkappa)^{-1} c_p g^{-1} p_{00}^{-\varkappa} \int_0^{\infty} \left(\overline{p^{1+\varkappa}} - \bar{p}^{1+\varkappa}\right) d\Theta \qquad (4)$$

where the bar over $p^{1+\varkappa}$ again denotes an average over an isentropic surface.

Since p is always positive and $1 + \varkappa > 1$, it is readily shown that $\overline{p^{1+\varkappa}} - \bar{p}^{1+\varkappa} > 0$ unless $p \equiv \bar{p}$. The precise magnitude of $\bar{A}$, particularly as compared to the average total potential energy per unit area, $\bar{P} + \bar{I}$, is less obvious. Expansion in a power series will aid the comparison. Thus, if $p = \bar{p} + p'$, it follows from the binomial theorem, applied to $p^{1+\varkappa}$, that

$$\overline{A} = (1+\varkappa)^{-1} c_p g^{-1} p_{00}^{-\varkappa} \int_0^\infty \overline{p}^{1+\varkappa} \cdot$$

$$\cdot \left[\frac{\varkappa(1+\varkappa)}{2!} \overline{\left(\frac{p'}{\overline{p}}\right)^2} - \right.$$

$$\left. - \frac{(1-\varkappa)\varkappa(1+\varkappa)}{3!} \overline{\left(\frac{p'}{\overline{p}}\right)^3} + \ldots \right] d\Theta \qquad (5)$$

The ratio of $\overline{A}$ to $\overline{P}+\overline{I}$ is a suitable mean value of the quantity enclosed in square brackets.

The power series (5) must converge if $p' < \overline{p}$ everywere, but the rapidity with which it converges depends upon typical values of $p'/\overline{p}$. The distribution of p' is in general far from normal, since tropospheric isentropic surfaces tend to be nearly horizontal in the tropics, so that p' is close to its maximum value over about half the area of the earth. Suppose that on a particular isentropic surface $p = 1{,}000$ mb over half the area, and p decreases linearly from 1,000 mb to 300 mb over the remaining half. In this case $\overline{p} = 825$ mb, $\overline{p'^2}/\overline{p}^2 = 0.075$, and $\overline{p'^3}/\overline{p}^3 = -0.019$. The ratio of the second to the first term in the power series is therefore—0.06, so that even in this rather extreme case, the power series is well represented by its leading term.

Therefore, approximately

$$\overline{A} = \frac{1}{2} \varkappa c_p g^{-1} p_{00}^{\varkappa} \int_0^\infty \overline{p}^{1+\varkappa} \overline{\left(\frac{p'}{\overline{p}}\right)^2} d\Theta \qquad (6)$$

and $\overline{A}$ depends upon the *variance* of pressure over the isentropic surfaces.

This variance is closely related to the variance of potential temperature on an isobaric surface, which in turn resembles the variance of temperature on an isobaric or horizontal surface. If $\overline{\Theta}$ and $\overline{T}$ are the average values of Θ and T on a isobaric surface, and Θ' and T' are the departures of Θ and T from $\overline{\Theta}$ and $\overline{T}$, the function $\overline{\Theta}(p)$ is not completely determined by the function $\overline{p}(\Theta)$, but approximately $p = \overline{p}(\overline{\Theta}(p))$, so that

$$p' = \overline{p}(\Theta - \Theta') - \overline{p}(\Theta) \sim -\Theta' \partial\overline{p}/\partial\Theta \qquad (7)$$

Thus

$$\overline{A} = \frac{1}{2} \varkappa c_p g^{-1} p_{00}^{-\varkappa} \int_0^{p_0} \overline{\Theta}^2 p^{-(1-\varkappa)} \left(-\frac{\partial\overline{\Theta}}{\partial p}\right)^{-1} \cdot$$

$$\cdot \overline{\left(\frac{\Theta'}{\overline{\Theta}}\right)^2} dp \qquad (8)$$

From the hydrostatic equation, it follows that

$$\partial\Theta/\partial p = -\varkappa\Theta p^{-1}(\Gamma_d - \Gamma)\Gamma_d^{-1} \qquad (9)$$

where $\Gamma = -\partial T/\partial z$ is the lapse rate of temperature and $\Gamma_d = g c_p^{-1}$ is the dry-adiabatic lapse rate. Since $\Theta'/\Theta = T'/T$,

$$\overline{A} = \frac{1}{2} \int_0^\infty \overline{T}(\Gamma_d - \overline{\Gamma})^{-1} \overline{\left(\frac{T'}{\overline{T}}\right)^2} dp \qquad (10)$$

Expression (10) is suitable for estimating the ratio $A/(P+I)$. This ratio evidently equals suitable average value of $\frac{1}{2}\Gamma_d(\Gamma_d - \Gamma)^{-1}\overline{T'^2}\overline{T}^{-2}$. The maximum values of $(\Gamma_d - \overline{\Gamma})^{-1}$ and probably also of $\overline{T'^2}$ occur in the troposphere. If $\overline{\Gamma} = \frac{2}{3}\Gamma_d$ and $\overline{T'^2} = (15°)^2$ are taken as typical values,

$$\overline{A}/(\overline{P} + \overline{I}) \sim 1/200$$

Hence less than one per cent of the total potential energy is generally available for conversion into kinetic energy.

3. Available potential energy and kinetic energy

It is a familiar observation that the total potential energy of the atmosphere greatly exceeds the kinetic energy. In considering the possible release of kinetic energy, however, we should compare the kinetic energy with the available potential energy.

The average kinetic energy per unit area of the earth's surface is approximately

$$\overline{K} = \frac{1}{2} g^{-1} \int_0^{p_0} \overline{V^2} dp \qquad (11)$$

In this expression we have neglected horizontal variations of p_0. From (2) it follows that

$$\overline{P} + \overline{I} = \frac{1}{\lambda - 1} g^{-1} \int_0^{\overline{p_0}} \overline{c^2}\, dp \qquad (12)$$

where $c^2 = \lambda RT$ is the square of the speed of sound. If we assume that a typical average wind speed is $^1/_{20}$ of the speed of sound, which lies between 300 m sec^{-1} and 350 m sec^{-1}, we find that

$$\overline{K}/(\overline{P} + \overline{I}) \sim 1/2000$$

If we also assume that our result

$$\overline{A}/(\overline{P} + \overline{I}) \sim 1/200$$

is typical, we find that

$$\overline{K}/\overline{A} \sim 1/10$$

Evidently, if kinetic energy is not released, it is not because a supply of available potential energy is lacking.

Let us see next how $\overline{K}$ and $\overline{A}$ vary. From the equation of continuity

$$\nabla \cdot \mathbf{v} + \partial\omega/\partial p = 0$$

where $\omega = \dot{p} = dp/dt$ is the individual pressure change, determined in the free atmosphere primarily by the vertical speed, and the thermodynamic equation

$$\partial\Theta/\partial t + \mathbf{v} \cdot \nabla\Theta + \omega\partial\Theta/\partial p = c_p^{-1}\Theta T^{-1}Q \quad (14)$$

where Q is the rate of addition of heat, per unit mass, we find that

$$\frac{1}{2}\partial\overline{\Theta'^2}/\partial t = -\overline{\Theta\omega}\,\partial\overline{\Theta}/\partial p - \frac{1}{2}\partial\overline{\Theta'^2\omega}/\partial p + \\ + C_p^{-1}\overline{\Theta}\,\overline{T}^{-1}\overline{\Theta'Q'} \qquad (15)$$

The second term on the right of (15) involves the space average of the product of three quantities, each of which is itself a departure from a space average. Such "triple correlations" are often negligible. In this case the term arises because another triple correlation, namely the term involving $\overline{p'^3}$ in (5), has been omitted in deriving expressions (6), (8),and (10) for $\overline{A}$ from (5). If we neglect the term involving $\overline{\Theta'^2\omega}$, we find, since $\Theta'/\overline{\Theta} = T'/\overline{T}$, that

$$\partial\overline{A}/\partial t = -C + G \qquad (16)$$

where

$$C = -Rg^{-1}\int_0^{\overline{p_0}} p^{-1}\,\overline{T\omega}\, dp = -\int_0^{\overline{p_0}} \overline{V \cdot \nabla z}\, dp \quad (17)$$

and

$$G = g^{-1}\int_0^{\overline{p_0}} \Gamma_d(\Gamma_d - \overline{\Gamma})^{-1}\overline{T}^{-1}\overline{T'Q'}\, dp \quad (18)$$

The latter integral in (17) is obtained from the former through the hydrostatic equation and the equation of continuity.

Likewise, from the equation of continuity and the equations of horizontal motion

$$\partial\mathbf{v}/\partial t + (\mathbf{v} \cdot \nabla)\mathbf{v} + \omega\,\partial\mathbf{v}/\partial p = \\ = 2\,\Omega \cdot \mathbf{v} - g\nabla z + \mathbf{F} \qquad (19)$$

where Ω is the vector angular velocity of the earth, and F is the horizontal force of friction, per unit mass, we find that

$$\partial\overline{K}/\partial t = C - D \qquad (20)$$

where

$$D = -g^{-1}\int_0^{\overline{p_0}} \overline{\mathbf{v} \cdot \mathbf{F}}\, dp \qquad (21)$$

Under adiabatic frictionless flow the *generation* G and the *dissipation* D vanish, so that the sum of $\overline{A}$ and $\overline{K}$ is conserved.

We have seen that the available potential energy depends upon the departure of the density stratification from horizontal. If the wind were exactly geostrophic everywhere, the kinetic energy of the whole atmosphere would be zero if and only if the available potential energy were zero. Since the actual wind tends to be nearly geostrophic throughout much of the atmosphere, it still follows that the kinetic energy is generally small or large according to whether the available potential energy is small or large. Large increases in available potential energy and kinetic energy should in general accompany each other.

We have seen, however, that under adiabatic flow increases in available potential energy and decreases in kinetic energy must accompany each other. It follows that when both forms of energy increase together, non-adiabatic effects are involved. Likewise, since

$\overline{K}$ is usually about one tenth of $\overline{A}$, any increase in $\overline{A}$ by more than about ten percent must involve nonadiabatic processes. Such processes are certainly responsible for the large increase in $\overline{A}$ from summer to winter, which is accompanied by a proportionately large increase in $\overline{K}$. This situation, which involves a decrease in $\overline{P}+\overline{I}$, has been discussed by Spar (1949).

We may ask at this point whether any appreciable changes in the ratio $\overline{K}/\overline{A}$ are possible under adiabatic flow, if the winds remain nearly geostrophic. That such changes are possible may be seen from the approximate expressions

$$\overline{A} = \frac{1}{2} K^{-1} g R^{-1} \int_0^{\overline{p_0}} \overline{\Gamma_d (\Gamma_d - \overline{\Gamma})^{-1} \overline{T}^{-1}} \cdot \overline{p^2 (\partial z'/\partial p)^2}\, dp \qquad (22)$$

$$\overline{K} \sim \frac{1}{2} g \int_0^{\overline{p_0}} \overline{f^{-2} (\nabla z)^2}\, dp \qquad (23)$$

obtained from (10) and (11) by substituting the hydrostatic and geostrophic relations, f being the Coriolis parameter. Both $\overline{A}$ and $\overline{K}$ depend upon the distribution of z' throughout the atmosphere, but, for a given variance of z', $\overline{K}$ is larger the more the fluctuations of z' in the horizontal, while $\overline{A}$ is larger the more the fluctuations of z' in the vertical. If the vertical variance spectrum of z' is relatively constant, the ratio $\overline{K}/\overline{A}$ will be larger when shorter wave lengths predominate in the horizontal variance spectrum of z'. It is thus possible for $\overline{K}$ and $\overline{A}$ to vary under adiabatic quasi-geostrophic flow. A familiar hypothetical example is the increase of kinetic energy which accompanies the exponential growth of short-wave pertubations superposed upon an unstable zonal current.

4. Zonal and eddy energy

An approach to the general circulation which has currently found much favor consists of resolving the field of motion into the mean zonal motion and the eddies superposed upon it. This resolution partitions the kinetic energy of the whole atmosphere into two types, which may be called *zonal kinetic energy* and *eddy kinetic energy*, and which represent the kinetic energies of the two types of motion. The maintenence of each type of kinetic energy is then considered. In addition to other forms of energy, each type of kinetic energy is a possible source or sink for the other type.

This partitioning of kinetic energy is essentially an analysis of variance of the wind field, and is possible because, aside from the contribution of the average wind, kinetic energy is the sum of the variances of the wind components. A similar analysis of variance of the temperature field is possible. To the extent that available potential energy is measured by the variance of temperature, this analysis partitions the available potential energy into two types, one due to the variance of zonally averaged temperature, and one due to the variance of temperature within latitude circles. These types may be called *zonal available potential energy* and *eddy available potential energy*. In addition to other forms of energy, each type of available potential energy may be a source or a sink for the other type.

Analytic expressions for $\overline{A_Z}$ and $\overline{A_E}$, the zonal and eddy available potential energies, and $\overline{K_Z}$ and $\overline{K_E}$, the zonal and eddy kinetic energies, per unit area, may be obtained from expressions (11) and (12) for $\overline{A}$ and $\overline{K}$ by replacing T and V by their zonal averages, or their departures from their zonal averages. Thus

$$\left.\begin{aligned} \overline{A}_Z &= \frac{1}{2} \int_0^{\overline{p_0}} (\Gamma_d - \overline{\Gamma})^{-1} \overline{T}^{-1} \overline{[T]'^2}\, dp \\ \overline{A}_E &= \frac{1}{2} \int_0^{\overline{p_0}} (\Gamma_d - \overline{\Gamma})^{-1} \overline{T}^{-1} \overline{T^{\star 2}}\, dp \\ \overline{K}_Z &= \frac{1}{2} g^{-1} \int_0^{\overline{p_0}} \overline{[V]^2}\, dp \\ \overline{K}_E &= \frac{1}{2} g^{-1} \int_0^{\overline{p_0}} \overline{V^{\star 2}}\, dp \end{aligned}\right\} \qquad (24)$$

Here square brackets denote an average with respect to longitude, at constant latitude and constant pressure, while an asterisk denotes

a departure from an average denoted by square brackets.

The time derivatives of these quantities may by applying suitable averaging processes to the continuity equation (13), the thermodynamic equation (14), and the equations of motion (19). Thus, if we allow the same sort of approximations which we used in the expressions for $\partial\overline{A}/\partial t$ and $\partial\overline{K}/\partial t$,

$$\left.\begin{aligned}
\partial\overline{A}_Z/\partial t &= -C_Z - C_A + G_Z \\
\partial\overline{A}_E/\partial t &= -C_E + C_A + G_E \\
\partial\overline{K}_Z/\partial t &= \;\; C_Z - C_K - D_Z \\
\partial\overline{K}_E/\partial t &= \;\; C_E + C_K - D_E
\end{aligned}\right\} \tag{25}$$

where

$$\left.\begin{aligned}
C_Z &= -Rg^{-1}\int_0^{\overline{p_0}} P^{-1}\overline{[T]}\,\overline{[\omega]}\,dp = \\
&= -\int_0^{\overline{p_0}} \overline{[\mathbf{v}]}\cdot\nabla\,\overline{[Z]}\,dp \\
C_E &= -Rg^{-1}\int_0^{\overline{p_0}} P^{-1}\overline{T^\star\omega^\star}\,dp = \\
&= -\int_0^{\overline{p_0}} \overline{\mathbf{v}^\star}\cdot\nabla\,\overline{Z^\star}\,dp \\
C_A &= -\frac{R}{g}\int_0^{\overline{p_0}} \frac{\overline{\Theta}}{\overline{T}}\left(\overline{[T^\star v^\star]}\frac{\partial}{\partial y}+\overline{[T^\star\omega^\star]'\frac{\partial}{\partial p}}\right)\cdot \\
&\qquad \cdot\left(\frac{\Gamma_d}{\Gamma_d-\overline{\Gamma}}\frac{1}{\overline{\Theta}}\overline{[T]'}\right)dp \\
C_K &= -\frac{1}{g}\int_0^{\overline{p_0}} \cos\phi\left(\overline{[u^\star v^\star]}\frac{\partial}{\partial y}+\right. \\
&\qquad \left.+\overline{[u^\star\omega^\star]}\frac{\partial}{\partial p}\right)\left(\frac{\overline{[u]}}{\cos\phi}\right)dp \\
&\quad -\frac{1}{g}\int_0^{\overline{p_0}}\left(\overline{[v^{\star 2}]}\frac{\partial}{\partial y}-\sin\phi\,\overline{[V^{\star 2}]}+\right. \\
&\qquad \left.+\overline{[v^\star\omega^\star]}\frac{\partial}{\partial p}\right)\overline{[v]}\,dp
\end{aligned}\right\} \tag{26}$$

$$\left.\begin{aligned}
G_Z &= g^{-1}\int_0^{\overline{p_0}} \Gamma_d(\Gamma_d-\overline{\Gamma})^{-1}\,\overline{T}^{-1}\overline{[T]'[Q]'}\,dp \\
G_E &= g^{-1}\int_0^{\overline{p_0}} \Gamma_d(\Gamma_d-\overline{\Gamma})^{-1}\,\overline{T}^{-1}\,\overline{T^\star Q^\star}\,dp \\
D_E &= -g^{-1}\int_0^{\overline{p_0}} \overline{[\mathbf{v}]\cdot[\mathbf{F}]}\,dp \\
D_Z &= -g^{-1}\int_0^{\overline{p_0}} \overline{V^\star\cdot F^\star}\,dp
\end{aligned}\right\} \tag{27}$$

In (26), u and v are the eastward and northward components of V, ϕ is latitude, and $\partial/\partial y$ is the derivative with respect to distance northward. The alternative forms for C_Z and C_E are analogous to the alternative forms (17) for C.

In equations (25) we observe that each of the quantities C_Z, C_E, C_A, and C_K occurs twice, with opposite signs. It is then tempting to say, for example, that C_K represents the rate of conversion from zonal to eddy kinetic energy, and to draw analogous conclusions about the other "C's", i.e., to interpret the C's as energy transformation functions, as described by Miller (1950). We must note, therefore, that the C's are not uniquely defined by the time derivatives of the various forms of energy, since, for example, if all the C's were altered by the additions of the same quantity, equations (25) would still be valid. To justify the interpretation of the C's as conversions from one form of energy to another, we must examine the physical processes which they describe. The necessity for considering physical processes when interpreting energy equations has recently been emphasized by Lettau (1954).

We note first the quantities G_Z and G_E, which may be called the *zonal generation* and the *eddy generation*, represent the generation (or destruction) of available potential energy by nonadiabatic processes, and do not involve conversion from one form of atmospheric energy to another. Similarly, the quantities D_Z and D_E, which may be called the *zonal dissipation* and the *eddy dissipation*, represent the dissipation of kinetic energy by friction,

and do not involve conversions of energy. Here we have regarded friction as involving a simultaneous destruction of kinetic energy and generation of potential energy, rather than a process of conversion from kinetic to potential energy, since very little available potential energy is generated by frictional heating.

It follows that the sum of the C's in the change of any one form of energy must equal the sum of the conversions to that form of energy from all other forms.

We next note that the C's all involve horizontal or vertical transports of momentum or sensible heat. These transports may be resolved into separate modes of transport; for example, the vertical transport of sensible heat, represented by $[T\omega]$, may be resolved into a transport by meridional circulations, an eddy-transport whose value per unit area is independent of latitude, and an eddy-transport whose value vanishes when averaged over latitude. Each of these modes of transport enters only one of the terms in the relation

$$[T\omega] = [T][\omega] + [T^{\star}\omega^{\star}] + [T^{\star}\omega^{\star}]' \quad (28)$$

Let us agree to regard the separate modes of transport as separate physical processes.

We then observe the following situation: Horizontal eddy-transports of sensible heat, and vertical eddy-transports whose values vanish when averaged over latitude, enter the expression for C_A, but not C_K, C_Z, nor C_E. They therefore affect A_Z and A_E by altering the analysis of variance of temperature, but they do not affect K_Z, K_E, nor the sum $A_Z + A_E$.

Eddy-transports of momentum enter the expression for C_K, but not C_A, C_Z, nor C_E. They therefore affect K_Z and K_E by altering the analysis of variance of wind, but they do not affect A_Z, A_E, nor the sum $K_Z + K_E$.

Transports of sensible heat by meridional circulations, and accelerations due to horizontal displacements by meridional circulations, enter the equivalent expressions for C_Z, but not C_E, C_A, nor C_K. They therefore affect A_Z and K_Z by altering the variance of zonally averaged temperature and wind, but they do not affect A_E, K_E, nor the sum $A_Z + K_Z$.

Vertical eddy-transports of sensible heat whose value per unit area is independent of latitude, and accelerations due to horizontal displacements by eddies, enter the equivalent expressions for C_E, but not C_Z, C_A, nor C_K. They therefore affect A_E and K_E by altering the variance of temperature and wind within latitude circles, but they do not affect A_Z, K_Z, nor the sum $A_E + K_E$.

It follows that C_A, C_K, C_Z, and C_E are energy transformation functions, which involve respectively only available potential energy, only kinetic energy, only zonal forms of energy, and only eddy forms of energy.

The energy transformation function C_K has appeared frequently in recent works. It is a modification of an expression derived by Reynolds (1894) in connection with turbulent flow. It has been presented in nearly the same form by Van Mieghem (1952), while the first integral in expression (25) for C_K, which is the dominating term, has been discussed by Kuo (1951) and Starr (1953).

The energy transformation function C_A bears nearly the same relation to temperature which C_K bears to wind. It depends upon the transport of sensible heat along the gradient of temperature in much the same way in which C_K depends upon the transport of angular momentum along the gradient of angular velocity.

The two possible remaining energy transformation functions—the conversions from A_Z to K_E and from A_E to K_Z, do not enter equations (25). Moreover, if we regard the separate modes of transport as separate physical processes, there is no process which affects both A_Z, and K_E, or both A_E and K_Z. These remaining energy transformation functions therefore vanish identically.

It must be remembered that this conclusion depends upon our regarding the separate modes of transport as separate physical processes. Without the distinction between the two modes of eddy transport, it would be impossible to say whether or not a direct conversion of zonal available potential energy to eddy kinetic energy is possible, although conversion from any form of available potential energy to eddy kinetic energy, which involves nonvanishing values of $T^{\star}$, would still require the *presence* of eddy available potential energy. The distinction between the modes of eddy-transport is probably as logical, if not as familiar, as the

distinction between eddy transport and transport by meridional circulations. Without the latter distinction none of the energy transformation functions involving zonal or eddy available potential energy could be defined.

5. The maintenance of the energy of the general circulation

The zonal winds and the superposed eddy motions are not identical with the meridional pressure gradient and the superposed pressure perturbations, since the former are features of the distribution of momentum, which possesses kinetic energy, and the latter are features of the distribution of mass, which possesses available potential energy. It is therefore a legitimate problem to study exchanges of kinetic energy between the zonal winds and the eddies, without considering similar exchanges of available potential energy.

Nevertheless the zonal winds are often identified with the meridional pressure gradient, and cyclonic and anticyclonic circulations are often identified with the low and high pressure systems which almost always accompany them. Indeed the wind systems could not long maintain their identities without the accompanying pressure systems, and vice versa. It may therefore be possible to achieve a better understanding of the general circulation by regarding exchanges of kinetic energy and exchanges of available potential energy as features of a single problem.

The conversion C_K of zonal to eddy kinetic energy depends primarily upon the transport of angular momentum horizontally and vertically by eddies along the gradient of angular velocity. Recent computations by STARR (1953), based upon wind observations over the northern hemisphere (see STARR and WHITE 1954) confirm the earlier suspicions of some meteorologists that the horizontal transport is predominantly against the gradient of angular velocity, and yield the approximate value—10×10^{20} ergs per second for the integral of C_K over the northern hemisphere, so that the eddies appear to supply sufficient kinetic energy to the zonal flow to maintain it against frictional dissipation.

At this point we must become more specific and state just what sort of dissipation we are considering. The dissipation of zonal kinetic energy by molecular friction is probably very small. The principal "frictional" dissipation of zonal kinetic energy is instead due to small-scale turbulent eddies, and it is principally the kinetic energy of these eddies which is dissipated by molecular friction. It is therefore not correct to say that the eddies on the whole supply kinetic energy to the zonal flow, if all scales of eddies are included.

Instead, we must make a distinction between large-scale eddies, which, roughly speaking, are the eddies large enough to appear on synoptic weather maps, and the remaining small-scale eddies. It is then correct to say that the large-scale eddies supply enough kinetic energy to the zonal flow to maintain it against the dissipative effects of small-scale eddies. This important feature of the general circulation would be obscured if eddies of all scales were included in a single category.

Again, if only molecular friction and conduction were considered, the skin friction and surface heating would have nearly infinite values, per unit mass, throughout nearly infinitesimal depths. The generation and dissipation functions G_Z and D_Z would then depend largely upon the usually unmeasured temperatures and winds in a thin layer next to the ground. This difficulty is overovercome if eddy viscosity and conductivity replace molecular viscosity and conductivity, so that the skin friction and surface heating have moderate values throughout moderate depths. Let us agree, therefore, to regard only the large-scale eddies as eddies, and to include the small-scale eddies in a category with molecular motions.

The conversion C_A of zonal to eddy available potential energy depends primarily upon the transport of sensible heat horizontally and vertically across the gradient of temperature T'. The studies of STARR and WHITE (1954) confirm the generally accepted idea that the horizontal transport is with the temperature gradient, and computations based upon the results of this study yield the approximate value 200×10^{20} ergs per second for the integral of C_A over the northern hemisphere, so that C_A is about twenty times as large as C_K.

It follows that if the zonal winds and the meridional pressure gradient are regarded as separate manifestations of the same zonal pattern, and if the wind and pressure variations

within latitude circles are regarded as separate manifestations of the same eddies, it is not possible to say that the eddies maintain the zonal circulation. All that can be said is that the zonal pressure field maintains the eddy pressure variations, but the eddy motion maintains the zonal motion.

To understand the maintenance of the energy of the general circulation, it is therefore not sufficient to know the exchange of energy between the zonal circulation and the eddies. A knowledge of all the energy transformation functions and the generation and dissipation functions is required.

The presence of net heating in low latitudes and net cooling in high latitudes is a familiar feature of the general circulation; it leads to a generally positive value of $\overline{[T]'[Q]'}$, so that the zonal generation G_Z is positive, and indeed seems to represent the primary source of the energy of the general circulation. Crude estimates of G_Z, based upon radiation-balance figures of Albrecht (see HAURWITZ 1941), and neglecting the release of latent energy, yield the approximate figure $G_Z = 200 \times 10^{20}$ ergs per second so that G_Z and C_A are about equal.

Less obvious is the sign of the eddy generation G_E, which depends upon $\overline{T^\star Q^\star}$, and hence upon the correlation between temperature and heating within latitude circles. Presumably it is negative, in view of the probable warming of cold air masses and cooling of warm air masses in middle latitudes, but the possible preference of warm longitudes for the release of latent energy may suppress this negative value.

The dissipation functions D_Z and D_E may safely be regarded as positive. We have just seen that C_K is negative, while C_A is positive. It follows by continuity that C_E must be positive, since it represents the only remaining source for eddy kinetic energy. This positive value must be associated with sinking of colder air and rising of warmer air at the same latitude. Hemispheric data for the direct computation of C_E are unfortunately not available.

The sign of C_Z cannot be inferred by continuity, but the magnitude seems to be small, in view of the failure of hemispheric wind observations to reveal strong meridional cells. A value of -2×10^{20} ergs per second for the integral of C_Z over the northern hemisphere has been estimated by STARR (1954) from the data available. The negative sign occurs because the middle-latitude indirect cell occupies the zone of maximum temperature gradient.

We are thus led to the following picture of the maintenance of the energy of the general circulation: The net heating of the atmosphere by its environment in low latitudes and the net cooling in high latitudes result in a continual generation of *zonal available potential energy*. Virtually all of this energy is converted into *eddy available potential energy* by the eddies. Some of this energy may be dissipated through heating of the colder portions of the eddies and cooling of the warmer portions; the remainder is converted into *eddy kinetic energy* by sinking of the colder portions of the eddies and rising of the warmer portions. Some of this energy is dissipated by friction; the remainder is converted into *zonal kinetic energy* by the eddies. Most of this energy is dissipated by friction; a small residual is converted into zonal available potential energy again by an indirect meridional circulation.

In conclusion, let us see how our picture of the energy transformation compares with earlier descriptions of the general circulation. Certainly it bears little resemblance to any theory which attributes the conversion of potential into kinetic energy to a general rising motion in low latitudes and sinking in high latitudes. However, discussions of some of the alternative processes which we have described have been appearing with increasing frequency in recent meteorological literature.

The idea that eddy kinetic energy is the immediate source of zonal kinetic energy is closely related to the idea expressed by ROSSBY (1947, 1949) that large-scale mixing processes (eddies) may account for the distribution of zonal winds. EADY (1950) has described the inequality of the mean zonal angular velocity as a result of turbulence (eddies). Computations of a positive rate of conversion of eddy into zonal kinetic energy, based upon wind observations, have been presented by KUO (1951) and STARR (1953).

An idealized quantitative model in which

the motions associated with waves in the westerlies (eddies) are responsible for the generation of kinetic energy has been presented by MINTZ (1947). The importance of sinking cold air masses and rising warm air masses was pointed out by ROSSBY (1949). VAN MIEGHEM (1952) describes the conversion of potential energy into *eddy* kinetic energy as one of the two most important energy transformations on the scale of the general circulation, the other being the conversion of eddy into zonal kinetic energy.

The importance of a poleward eddy-transport of sensible heat, in conjunction with the excess of radiational heating in low latitudes and the deficit in high latitudes, has long been recognized (cf. HAURWITZ 1941), but the inevitable other effect of this transport—an increasing of the variance of temperature within latitude circles—seems to have been generally overlooked. Very recently STARR (1954) has described an idealized two-stage process of conversion of potential into kinetic energy, the first step consisting of a deformation of zonally oriented isotherms. This stage is essentially a conversion from zonal into eddy available potential energy. But regardless of whether one wishes to think in terms of available potential energy, a variance of temperature within latitude circles is a prerequisite for the conversion of potential energy into eddy kinetic energy, which involves a correlation within latitude circles between temperature and vertical motion (cf. VAN MIEGHEM 1952). If non-adiabatic heating creates primarily a cross-latitude variance of temperature, a poleward eddy-transport of sensible heat is necessary to maintain the variance of temperature within latitude circles. The conversion of zonal to eddy available potential energy may therefore be regarded as a third important energy transformation on the scale of the general circulation.

REFERENCES

EADY, E. T., 1950: The cause of the general circulation of the atmosphere. *Cent. Proc. Roy. Met. Soc.*, 156—172.

HAURWITZ, B., 1941: *Dynamic Meteorology*. New York, McGraw-Hill Book Co., 365 pp.

KUO, H. L., 1951: A note on the kinetic energy balance of the zonal wind systems, *Tellus*, **3**, 205—207.

LETTAU, H., 1954: Notes on the transformation of mechanical energy from and to eddying motion. *J. Meteor.*, **11**, 196—201.

MARGULES, M., 1903: Über die Energie der Stürme. *Jahrb. kais-kön Zent. für Met.*, Vienna. Translation by C. Abbe in *Smithson. Misc. Coll.*, **51**. 1910.

MILLER, J. E., 1950: Energy transformation functions. *J. Meteor.*, **7**, 152—159.

MINTZ, Y., 1947: On the kinematics and thermodynamics of the general circulation of the atmosphere in the higher latitudes. *Trans. Am. Geophys. U.*, **28**, 539—544.

REYNOLDS, O., 1894: On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Phil. Trans. Roy. Soc. (A)*, **186**, 123—164.

ROSSBY, C.-G., 1947: On the distribution of angular velocity in gaseous envelopes under the influence of large-scale horizontal processes. *Bull. Am. Met. Soc.*, **28**, 53—68.

— 1949: On a mechanism for the release of potential energy in the atmosphere. *J. Meteor.*, **6**, 163—180.

SPAR, J., 1949. Energy changes in the mean atmosphere. *J. Meteor*, **6**, 411—415.

STARR, V. P., 1953: Note concerning the nature of the large-scale eddies in the atmosphere. *Tellus*, **5**, 494—498.

— 1954: Commentaries concerning research on the general circulation. *Tellus*, **6**, 268—272.

STARR, V. P., and WHITE, R. M., 1954: *Balance requirements of the general circulation*. Geophysical research papers, Cambridge, Mass. (in press).

VAN MIEGHEM, J., 1952: Energy conversions in the atmosphere on the scale of the general circulation. *Tellus*, **4**, 334—351.

Instability Criteria and Growth of Baroclinic Disturbances

By ROBERT G. FLEAGLE, University of Washington[1]

(Manuscript received June 16, 1954, revised version received March 30, 1955.)

Abstract

The instability criterion and the rate of growth of long baroclinic disturbances in horizontally uniform zonal flow is derived for a sloping streamsurface. Both quantities are expressed by one explicit and simple function. Unstable disturbances are characterized by upward motion toward the cold air and downward motion toward the warm air. The slope of the streamsurface is eliminated, and the growth equation is compared with the results of other studies. The maximum rate of growth is small enough to indicate that numerical prediction of the development of baroclinic disturbances is practicable. Other results are that long waves may occur in a baroclinic atmosphere above a flat earth as well as above a spherical earth and that on a spherical earth unstable disturbances occur only north of a critical latitude.

1. Historical Perspective

The fundamental importance of differential heating in creation of the large scale atmospheric motions was stated long ago by HADLEY (1735—36). But only in recent years have observations been available to show that the actual motions differ markedly from the simple convective circulation envisaged by Hadley. Theory has been, if anything, even slower; the hydrodynamics of planetary atmospheres has been developed only in the past two decades, and it is still far from complete.

Observations in low latitudes show that the zonal speed increases with latitude in rough conformity with conservation of angular momentum, and the dominant motion is relatively slow and steady (UNIVERSITY OF CHICAGO, 1947). By contrast, in middle and high latitudes the atmosphere is subject to large amplitude disturbances, and the zonal speed decreases with latitude. These features of the atmospheric circulation pose several problems; viz., (1) why are the regimes of low and high latitude different? (2) how is the middle latitude maximum of zonal wind speed maintained? (3) under what conditions does this zonal current "break down" into large scale meridional currents? Clearly, these problems are interrelated, but their complexity makes it desirable to treat them separately. This paper is concerned mainly with the third problem and to a lesser extent with the first.

Observations suggest that the transformation from zonal to meridional currents is an instability phenomenon; that is, one in which small displacements are amplified with time. The standard method for determining the conditions necessary for instability (THOMSON, 1871) consists of discovering the conditions under which the speed of periodic disturbances in the fluid is a complex function. The imaginary part of the complex function yields an exponential increase of the amplitude with time.

[1] Contribution No. 13, Department of Meteorology and Climatology, University of Washington. Supported in part by the Geophysical Research Directorate, A. F. Cambridge Research Center.

Studies of instability in the atmosphere stem from Lord Rayleigh's (1878, 1880, 1887) investigations of the conditions under which small displacements at the boundary between two air streams grow into larger displacements and from Helmholtz' (1889) determination that unstable gravitational waves may form on a surface separating two layers of different density and velocity. These two discoveries led to further studies of the effects of velocity shear at the boundary between fluids, and the term *shearing instability* has been used to describe instability of this type. This series of investigations culminated in the polar front theory of cyclone formation (V. Bjerknes et al., 1933). As a result of the widespread interest in this theory, attention was focused on the region near a boundary between two fluids rather than on the behavior of a single extensive fluid.

A second series of investigations has stemmed from Helmholtz' (1888) and Rayleigh's (1913) studies of the relation of vorticity distribution to instability. Solberg (1936) showed that disturbances within a zonal atmospheric current are unstable if the meridional velocity shear along isentropic surfaces exceeds the value of the Coriolis parameter. Subsequent studies by Ertel (1940), Kleinschmidt (1941) and others and a more complete criterion derived by Kuo (1949) represent attempts to explain the more or less chronic instability of mid-latitude zonal flow on this basis. Instability of this type often is called *dynamic instability*, a particularly unfortunate choice of terms because the label *dynamic* implies that other types of instability are unrelated to forces. Despite the unquestioned validity of the instability criterion, observations gradually have made it clear that the criterion is so severe that the instability of zonal flow cannot be explained primarily by the quasi-horizontal shear of the zonal wind.

2. Baroclinic Instability

The development of the hydrodynamics of planetary atmospheres began with Rossby's (1939) demonstration that large-scale disturbances may exist in a barotropic atmosphere. Sutcliffe (1947) and Charney (1947) extended the theory to a baroclinic atmosphere; their results showed that if the vertical shear of the undisturbed zonal current (here called baroclinity) exceeds a certain critical value which depends on wavelength and latitude small disturbances will grow in amplitude. According to this criterion, waves of less than a critical length are unstable. The importance of this discovery quickly was recognized, so that since 1947 primary attention has shifted from instability arising from quasi-horizontal shear or from shear at boundaries between homogeneous air masses to instability arising from vertical wind shear. The term *baroclinic instability* has come to be used to describe this sort of instability, and extensions of the theory have been made.

Several investigators, notably Fjørtoft (1950), Berson (1951), Kuo (1952), and Thompson (1953), have succeeded in extending the theory into the domain of unstable waves. The results of these studies show that maximum instability exists for waves of intermediate length. However, there is considerable difference in these results. Kuo finds short waves unstable, whereas the others find them to be stable. Fjørtoft attributes short-wave stability to non-geostrophic velocity at short wavelengths, the others do not. Various boundary conditions have been used, but their effects on the solutions are not clear.

3. Status of the Problem

The foregoing brief survey of the history of baroclinic instability serves to illuminate the very rapid progress which has been made toward the solution of the problem. It also reveals that at present the role of non-geostrophic wind in short waves is uncertain, and that understanding of the effects of various simplifying assumptions is incomplete. One must recognize, too, that the most complete solution to the problem (Kuo, 1952) was possible only with mathematical complexity so great as to defy clear understanding of the role played by the various parameters of the problem.

Finally, it is appropriate to point out that each of the instability criteria contain parameters like vertical wind shear, hydrostatic stability, and undisturbed zonal wind speed which are, in part at least, products of the instability phenomenon and which, therefore, may depend on other parameters of the prob-

lem. For this reason, it is possible that disturbances which are indicated as unstable by a theoretical criterion in reality may be stable, or vice-versa. To put it differently, arbitrary choice of the parameters of the problem may not be proper.

There is, therefore, only limited importance in a discussion of whether one instability criterion or another is superior. The problem of greatest interest and importance at the present time is to incorporate more of the parameters into the solution or at least to gain additional insight into the relation of parameters to solution. Eventually, of course, one hopes to consider a model in which the only arbitrary parameter is the distribution of solar energy absorbed at the earth's surface. Then the problem of the general circulation may be said to have been solved.

4. Purpose

The purpose of this paper is to clarify the influence of the various parameters in the development of long baroclinic disturbances, and, thereby, to gain further insight into the mechanism of the general circulation.

5. Method

The central difficulty encountered in developing a frequency equation for a compressible baroclinic atmosphere appears when the solution or solutions are substituted into the boundary conditions. Typically, the resulting equation is a complex function and is hideously transcendental. However, this difficulty may be avoided if it is possible to incorporate the boundary conditions or the results of observation in the equations at the outset in such a way that the number of dependent variables and their derivatives is equal to the number of independent equations. If this can be done, the frequency equation emerges promptly upon algebraic reduction.

The normal component of velocity is required to vanish at rigid impervious surfaces; therefore, immediately above the horizontal ground surface the streamlines are horizontal. Computations by FLEAGLE (1947) and others indicate that above the ground the streamlines slope upward toward the pole. In a broad zonal current these streamlines are approximately plane so that planes may be passed through individual streamlines to form surfaces as shown in Fig. 1. These surfaces are referred to hereafter as streamsurfaces. The maximum slope of the streamsurfaces is reached in the middle troposphere, not always at the same height. If the streamsurface of greatest slope is selected as the x, y plane of a Cartesian coordinate system, then the velocity normal to the x, y plane and the normal gradient of the normal velocity vanishes everywhere on this plane. It follows that, so long as the equations are applied to the streamsurface of maximum

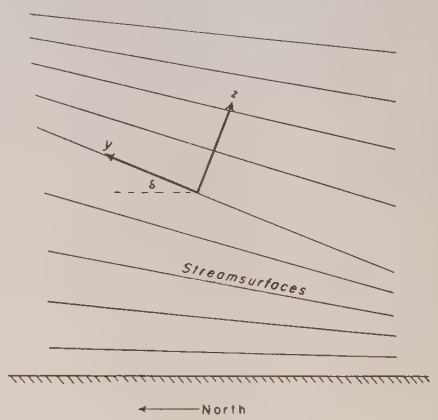


Fig. 1. Vertical cross section showing sloping streamsurfaces and orientation of the coordinate system.

slope, the vertical component of velocity and its space derivatives do not appear in the hydrodynamic equations.

Upon adopting the coordinate system described above, the five perturbation equations which describe a compressible baroclinic atmosphere contain the dependent variables: pressure, density, vertical pressure gradient, and the velocity components in the x and y directions.

The enormous simplification of the problem which is introduced by the elimination of the vertical velocity and its height derivative makes it possible to eliminate all five variables by algebraic reduction. This procedure, of course, leads only to the frequency equation; the vertical structure of the disturbance must be developed from the complete set of equations including the vertical velocity terms. Observations indicate that growth of disturbances occurs simultaneously and at about the same rate throughout a large part of the troposphere.

Therefore, although the frequency equation derived for the streamsurface of greatest slope applies directly only to that surface, it also applies indirectly to a layer of considerable thickness.

6. The Frequency Equation

In the sloping Cartesian coordinate system described above the vectors representing the acceleration of gravity and the angular velocity of the earth may be resolved into components as follows:

$$\mathbf{g} = -g\,(\mathbf{j}\sin\delta + \mathbf{k}\cos\delta) \tag{1}$$

$$\boldsymbol{\Omega} = \Omega\,(\mathbf{j}\cos(\phi-\delta) + \mathbf{k}\sin(\phi-\delta)) \tag{2}$$

The symbols used here and throughout the paper are defined in the Appendix.

The perturbation method as described by HAURWITZ (1951) is accurate for infinitesimal disturbances and therefore is the natural method for determination of instability criteria. The undisturbed motion will be assumed to be characterized by unchanging zonal flow which is constant in the x and y directions and which increases linearly with height (normal to the earth's surface). The equations of undisturbed motion given by HAURWITZ (1951) then become

$$V = 0 \tag{3}$$

$$fU = -Q^{-1}P_y - g\sin\delta \tag{4}$$

$$g\cos\delta = -Q^{-1}P_z \tag{5}$$

$$W_z = 0 \tag{6}$$

$$dQ/dt = 0 \tag{7}$$

From (4) and (5) it follows that

$$Q^{-1}Q_y = \frac{fU_z}{g\cos\delta} - g\Gamma\sin\delta - fU\Gamma \tag{8}$$

where $Q^{-1}Q_z$ is represented by $-g\Gamma\cos\delta$.

Because the undisturbed current is assumed to be independent of y, it seems reasonable to impose the arbitrary restriction that the velocity perturbations are independent of y. The only justification for this restriction is the observation that velocity disturbances which are approximately independent of y do occur, and the conditions under which they develop do not appear to be markedly different from the conditions under which velocity disturbances develop which are dependent on y. The perturbation equations given by HAURWITZ (1951) may be written for this case in the form

$$L(u) + \underline{wU_z} - fv = -Q^{-1}p_x \tag{9}$$

$$L(v) + fu = -Q^{-1}p_y + qQ^{-2}P_y \tag{10}$$

$$\underline{L(w)} + gqQ^{-1} = -Q^{-1}p_z \tag{11}$$

$$L(q) + vQ_y + Qu_x + \underline{(Qw)_z} = 0 \tag{12}$$

$$L(q) + vQ_y + \underline{wQ_z} - \gamma\,[L(p) + vP_y + \underline{wP_z}] = 0 \tag{13}$$

Upon applying these equations to the streamsurface of maximum slope, the underlined terms vanish. Then, if p_y is eliminated by cross differentiation of (9) and (10)

$$L(v_x) + fu_x + v\beta = Q^{-2}(q_xP_y - p_xQ_y) \tag{14}$$

where β represents $\partial f/\partial y$.

Equations (9)—(14) are linear equations so that solutions may be superimposed. Therefore, no loss of generality occurs if the disturbances are assumed to have the simple form given by

$$u = A(z)\,e^{i\alpha(x-ct)} \tag{15}$$

$$v = B(z)\,e^{i\alpha(x-ct)} \tag{16}$$

$$p = D(z)\,e^{i\alpha(x-ct)} \tag{17}$$

$$q = F(z)\,e^{i\alpha(x-ct)} \tag{18}$$

Substitution of (15)—(18) in (9), (11), (12), (13), (14) yields five simultaneous algebraic equations in A, B, D, D_z, and F. In order that non-trivial solutions exist for A, B, D, F, and D_z, the determinant of the coefficients must vanish. It follows directly that

$$-\gamma(c-U)^3[\alpha^2(c-U)+\beta] + (c-U)^2(\alpha^2 + f^2\gamma) + (c-U)[\beta + fQ^{-1}(Q_y - 2\gamma P_y)] - Q^{-2}P_y(Q_y - \gamma P_y) = 0 \tag{19}$$

The first term in (19) is small compared to the second for all values of $c-U$ much smaller

than the speed of sound; it therefore is neglected. If the meridional velocity had been assumed to obey the geostrophic equation, this term would not have appeared; but equation (19) would otherwise be unchanged. So, it must be concluded that the quasi-geostrophic assumption does not lead to significant error *for disturbances in which the velocity perturbations are independent of* y.

Elimination of the first term in (19) leads directly to

$$c - U = -\frac{1}{2(\alpha^2 + f^2\gamma)}\{\beta + fQ^{-1}(Q_y - 2\gamma P_y) \mp$$

$$\mp[\beta^2 + 2\beta f Q^{-1}(Q_y - 2\gamma P_y) + 4\alpha^2 Q^{-2} P_y (Q_y - \gamma P_y) + f^2 Q^{-2} Q_y^2]^{1/2}\} \qquad (20)$$

Substitution of (4) and (8) in (20) gives a cumbersome equation which, for long baroclinic disturbances, may be simplified by the following approximations.

$$\alpha^2 + f^2\gamma \approx \alpha^2 \qquad \sin\delta \approx \delta$$

$$2\gamma - \Gamma \approx \gamma \qquad \cos\delta \approx 1$$

$$U_z - 2Us_z \approx U_z$$

$$\beta + 2f^2 U\gamma + 2f^2 U_z/g \approx \beta$$

It follows that (20) may be written

$$c - U = -\frac{\beta + fg\gamma\delta}{2\alpha^2} \mp \frac{1}{\alpha}\left\{\frac{1}{4\alpha^2}(\beta + fg\gamma\delta)^2 - \right.$$

$$\left. - \delta(fU_z - gs_z\delta)\right\}^{1/2} \qquad (21)$$

For unstable disturbances the wave speed given by (21) is complex; the real part describes the speed and the imaginary part describes the rate of growth of the unstable waves. Thus

$$c_r - U = -\frac{\beta + fg\gamma\delta}{2\alpha^2} \qquad (22)$$

$$\alpha c_i = \left\{\delta(fU_z - gs_z\delta) - \frac{1}{4\alpha^2}(\beta + fg\gamma\delta)^2\right\}^{1/2} \qquad (23)$$

The preceding derivation may be performed in spherical coordinates as readily as in Cartesian coordinates. The result is very nearly identical with (22) and (23) except very near the poles. Equation (23) shows that instability depends in a direct sense on U_z, the baroclinity. However, instability is possible only if δ is positive; therefore, upward motion to the east and downward motion to the west of pressure troughs is essential in order that disturbances grow. This relationship is reflected in the fact that cloudiness and precipitation occur to the east of low pressure systems and clear skies prevail to the west. The first term in (23) indicates that instability is possible for positive $gs_z\delta$ only if $gs_z\delta$ is less than fU_z; that is, if the slope of the streamsurface is less than the slope of the isentropic surface. The second term in (23) indicates that very long waves are stable and shorter waves are unstable. This conclusion is valid, of course, only if δ, as well as other parameters, is reasonably independent of wavelength.

For the case of horizontal motion, δ vanishes; only stable disturbances are possible, and (21) reduces to the trough formula for a barotropic atmosphere. This reduction demonstrates that, although horizontal velocities in large disturbances are three orders of magnitude greater than the vertical velocities, vertical velocity plays a role of crucial importance in the conversion of the potential energy of the baroclinic atmosphere to kinetic energy.

7. Elimination of the Slope of the Streamsurface

The slope of the streamsurface may be eliminated from (21) by evaluating the ratio of the vertical to the meridional velocity components at the surface of its maximum value. This requires manipulation of the perturbation equations (9), (11), (12), (13), (14) including the vertical velocity terms. Upon eliminating from these equations the perturbation pressure and perturbation velocity in the x direction, the following equations result.

$$fL(q_x) - Q^{-1}Q_y L^2(q) + Q^{-1}P_y q_{xx} + Q_y U_z w_x -$$

$$- Q^{-1}Q_y^2 L(v) - QL(v_{xx}) - Q\beta v_x + f(Qw)_{zx} -$$

$$- Q^{-1}Q_y L[(Qw)_z] = 0 \qquad (24)$$

$$L(q_{xx}) - \gamma L^3(q) + Q_z w_{xx} + \gamma U_z QL(w_x) +$$

$$+ g\gamma Q w_{xx} + Q_y v_{xx} - f\gamma QL(v_x) - \gamma Q_y L^2(v) -$$

$$\gamma P_y v_{xx} - \gamma L^2[(Qw)_z] = 0 \qquad (25)$$

$$U_z L(q_x) + g\,q_{xx} + L^2(q_z) + Q_y L(v_z) + \\ + L[(Qw)_{zz}] + QL(w_{xx}) + U_z Q_y v_x + fQ_z v_x + \\ + fQv_{zx} = 0 \qquad (26)$$

Then, upon substituting solutions (15)—(18) there result three algebraic equations. Small terms may be eliminated at this point if care is taken to ensure that the remaining large terms do not cancel at a subsequent step. The relative magnitudes of all small terms but two are independent of wavelength. The exceptions are negligible for all disturbances more than several hundred kilometers in length. Elimination of the small terms leads to the equations

$$\alpha cfF + if(QC)_z - i[\alpha^2(c-U) + \beta]QB = 0 \qquad (27)$$

$$i\alpha(c-U)F + s_z QC - f[U_z/g + \gamma c - \\ \Gamma U]QB = 0 \qquad (28)$$

$$\alpha gF - if(QB)_z = 0 \qquad (29)$$

where C represents the amplitude of the vertical velocity perturbation. It is easy to show that if quasi-geostrophic and quasi-hydrostatic equilibrium had been assumed at the outset, equations (27), (28), and (29) would have resulted without the final elimination of small terms. So, *for disturbances in which the velocity perturbations are independent of the y coordinate*, these assumptions are valid, at least down to wavelengths of a few hundred kilometers.

If F is eliminated from (27), (28), and (29)

$$gf(QC)_z + f^2 c(QB)_z = g[\alpha^2(c-U) + \beta]QB \qquad (30)$$

$$gs_z C + f(cs_z - U_z)B = f(c-U)B_z \qquad (31)$$

At the streamsurface of greatest slope $(QC)_z = \delta(QB)_z$. This condition serves the function of boundary conditions in the complete integration of the hydrodynamic equations and makes it possible to eliminate $(QC)_z$ and $(QB)_z$ between (30) and (31). If, as before, only terms of the largest order of magnitude are retained, there results

$$\delta = \frac{f[U_z - g\Gamma(c-U)]}{gs_z} \qquad (32)$$

If only unstable disturbances are considered, $(c-U)$ may be eliminated from (32) by the use of (22). This gives

$$\delta = \frac{2fU_z\alpha^2 - \beta fg\Gamma}{2gs_z\alpha^2 + f^2g^2\gamma\Gamma} \qquad (33)$$

Substitution of (33) in (22) and (23) gives

$$c_r - U = -\frac{f^2\gamma U_z + \beta s_z}{2s_z\alpha^2 + f^2 g\gamma\Gamma} \qquad (34)$$

$$\alpha c_i = \frac{f\Gamma}{2s_z\alpha^2 + f^2 g\Gamma\gamma} \\ \{\alpha^2(f^2U_z^2 - \beta^2 s_z^2/f^2\Gamma^2) - \beta(gf^2\Gamma U_z + g\beta s_z)\}^{1/2} \qquad (35)$$

These relations can be obtained without the prior derivation of (21), but this would leave somewhat obscure the role of vertical motion in the release of baroclinic instability.

Inspection of (35) shows that short disturbances (large α) are slightly unstable, longer disturbances are more unstable, but very long disturbances are stable. The wavelength of greatest instability is given by

$$\lambda_m = 2\pi\left\{\frac{2s_z(f^2U_z^2 - \beta^2 s_z^2/f^2\Gamma^2)}{gf^2\Gamma U_z(f^2\Gamma U_z + 4\beta s_z) + 3g\beta^2 s_z^2}\right\}^{1/2} \qquad (36)$$

8. Comparison with Results of Other Studies

The critical curves and the amplification per 24 hours computed from (35) is shown in Fig. 2. The parameters chosen are those used by CHARNEY (1947), FJØRTOFT (1950), and KUO (1952). The critical curve found by Charney and Kuo is indicated for comparison. The critical curve for a different lapse rate also is shown.

The difference between the results shown in Fig. 2 and KUO's (1952) results must reflect differences in the atmospheric models used. The advantage of Kuo's treatment lies in its generality; his frequency equation was obtained from the perturbation equations by introducing special and rather simple boundary conditions. The advantage of the treatment given here lies in its simplicity; the result was obtained by using the fact of the existence of a streamsurface of maximum slope. It may be noted that Kuo's solution requires the slope of the streamsurfaces to increase upward without limit, presumably, as a result of an unrealistic upper boundary condition. It is

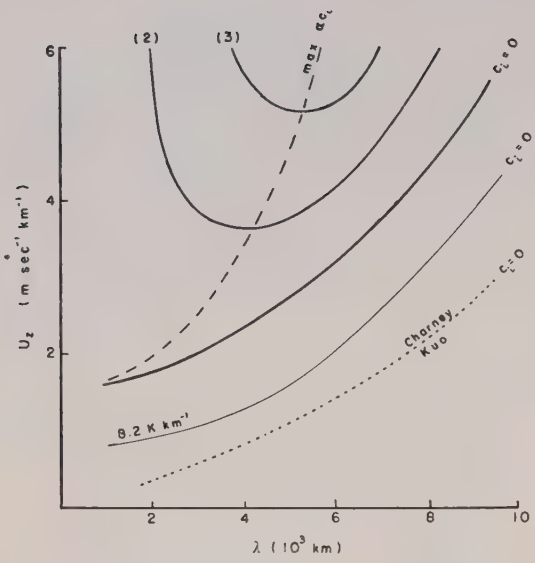


Fig. 2. Critical curves (labelled $c_i = 0$) and the 24 hour amplification (labelled (2) and (3)) computed from (35) for latitude of 45° N, and temperature lapse rate of 6.5 K km^{-1} compared with the critical curves found by CHARNEY (1947) and KUO (1952). The critical curve for a temperature lapse rate of 8.2 K km^{-1} is shown by the thin line.

not possible to conclude which frequency equation is the more accurate description of the real atmosphere, but this is a matter of little importance because other simplifications in the models and in the mathematical treatment probably exert greater influences on the final accuracy. One need only mention the assumption of horizontally uniform unchanging mean flow with infinitesimal perturbations, an assumption which is never fulfilled by the atmosphere.

In order to gain insight into the short-wave stability found by FJØRTOFT (1950), BERSON (1951), and THOMPSON (1953), the stream-surface may be eliminated from (22) and (23) by integrating (30) from sea level to the surface of greatest slope. If the height of this surface is chosen independently of wavelength, short waves are markedly stable. Similar arbitrary choices of characteristic heights were made by Fjørtoft, Berson, and Thompson with very similar results. This comparison indicates that the short-wave stability arose from the assumption that a characteristic height was independent of wavelength; the stability probably is not inherent in the real atmosphere.

9. The Rate of Growth

Fig. 2 indicates that the amplitude of disturbances cannot grow at a rate exceeding a factor of from one to three per day for the vertical shear and hydrostatic stability commonly observed. It is difficult, therefore, to believe that major disturbances ever develop by the process described here from very small more or less random disturbances; rather, the disturbance in its initial stage must be only about an order of magnitude less than in its fully developed stage. The initial disturbance presumably could be created by topography, by heating, or by the effects of dispersion acting on disturbances in a remote part of the zonal current. In any case, the fairly deliberate growth of these unstable disturbances makes prediction of their amplitude and movement a practicable possibility. This must be carried out, of course, by numerical integration of nonlinear equations. However, insights gained from this study of linear equations may be of value in the selection of nonlinear atmospheric models.

10. Effect of the Spherical Shape of the Earth

Equation (35) shows that for any combination of baroclinity and wavelength, instability increases with latitude. In Fig. 3 the critical curves are illustrated for several baroclinities.

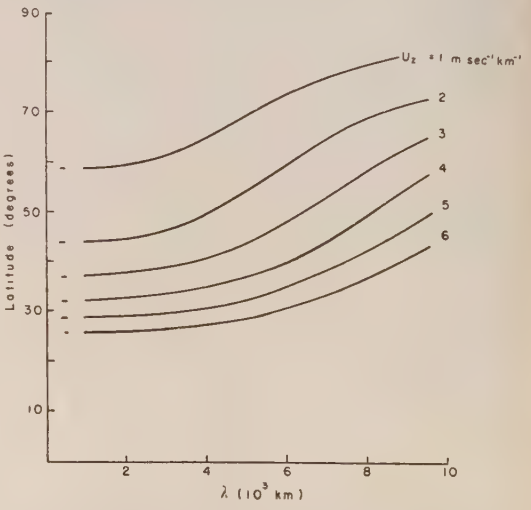


Fig. 3. Critical latitude south of which disturbances are stable for a temperature lapse rate of 6.5 K km^{-1} ((computed from (35)).

Instability is possible for reasonable baroclinities only north of about 30 degrees. This is consistent with the observation that large scale disturbances are characteristic of middle and high latitudes and do not penetrate to low latitudes.

On a flat earth, f is constant and β vanishes. Equation (34) shows that disturbances characteristic of observations are still possible, so long as the atmosphere is baroclinic. These considerations make understandable the observations of long waves in a rotating dishpan, described by Long.[2] The generation of the waves in this case is to be attributed to the baroclinity rather than to conservation of potential vorticity.

11. Conclusions and Inferences

The frequency equation for long baroclinic disturbances has been expressed in very simple form (equations 34 and 35). The results are in substantial agreement with those of Kuo (1952), but the simpler form derived here makes it easy to determine the dependence of the wave speed and the rate of growth of unstable disturbances on baroclinity, hydrostatic stability, wavelength, and latitude.

Equations (34) and (35) show that disturbances of about four thousand kilometer length may be expected to develop in middle latitudes where the baroclinity reaches 3 to 4 m sec^{-1} km^{-1}. On the other hand instability is not to be expected south of about 30° latitude. It follows that a relatively steady large scale circulation controlled primarily by differential heating and conservation of angular momentum is more likely in the tropics than at higher latitude.

The results suggest a qualitative insight into the problem of the meridional distribution of zonal wind speed. Although (35) indicates that greatest instability should occur at the pole, it is intuitively plausible that the atmosphere selects the mode of instability which permits the greatest conversion of potential to kinetic energy. This would occur sufficiently far from the pole that a substantial part of the atmosphere could take part in the instability process but still north of the critical latitude. Viewed in this way, the maximum zonal speed observed in middle latitudes is a consequence of the conversion of potential energy of the baroclinic atmosphere to kinetic energy. It may be analogous to the localized vertical convection currents associated with hydrostatic instability.

[2] Long, R. R., 1951: Research on experimental hydrodynamics in relation to large-scale meteorological phenomena. Prog. Rep. No. 5, 55 pp.

Appendix

List of Symbols:

g: acceleration of gravity
δ: slope of the streamsurface
ϕ: latitude
Ω: angular velocity of earth
U: undisturbed velocity component in x direction
u, v, w: perturbation velocity components in x, y and z directions, respectively
f: $2\,\Omega \sin(\phi - \delta)$ and $2\,\Omega \sin \phi$
t: time
Q, q: undisturbed and perturbation densities, respectively
P, p: undisturbed and perturbation pressures, respectively
j, k: unit vectors in y and z directions, respectively
A,B,C,D,F: amplitudes of u, v, w, p, q, perturbations, respectively
Subscripts: partial differentiation with respect to that variable
γ: coefficient of piezotropy
Γ: $-(gQ)^{-1} Q_z \sec \delta$ and $-(gQ)^{-1} Q_z$
L: operator, $(\partial/\partial t + U\,\partial/\partial x)$
β: $\partial f/\partial y$
α: wave number
λ: wavelength
c: complex wave speed
c_r: real part of wave speed of unstable wave
ic_i: imaginary part of wave speed
s_z: hydrostatic stability

REFERENCES

Berson, F. A., 1951: On the factors controlling the instability of long waves in zonal currents. *Archiv für Geophysik* **1**, 187—236.

Bjerknes, V., et al., 1933: *Physikalische Hydrodynamik*. J. Springer, Berlin, 797 pp.

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *J. Meteor.*, **4**, 135—162.

Ertel, H., 1940: Über die Stabilität der zonalen atmosphärishen Zirkulation. *Meteor. Zeit.*, **57**, 397—400.

Fjørtoft, R., 1950: Application of integral theorems in deriving criteria of instability for laminar flows and for the baroclinic vortex. *Geofys. Publ.*, **17**, 5, 52 pp.

Fleagle, R. G., 1947: The fields of temperature, pressure, and three-dimensional motion in selected weather situations, *J. Meteor.*, **4**, 165—185.

Hadley, G., 1735—36: Concerning the cause of the general trade wind. *Phil. Traus. Roy. Soc.*, London, **39**: 58—62.

Haurwitz, B., 1951: The perturbation equations in meteorology. *Comp. of Meteor.*, 401—420. Am. Meteor., Soc., Boston.

Helmholtz, H., 1888: Über atmosphärische Bewegungen. *Meteor. Zeit.*, **5**, 329—340.

— 1889: On atmospheric movements (2nd paper) *Smithsonian Misc. Coll.* **34** (1893), 94—111 (Translated by C. Abbe from *Sitzungs-Berichte, Roy. Pruss. Acad. Sci.*, Berlin, 1889, 761—780).

Kleinschmidt, E., 1941: Zur Theorie der labilen Anordnung. *Meteor. Zeit.*, **58**, 157—163.

Kuo, H. L., 1949: Dynamic instability of two-dimensional nondivergent flow in a barotropic atmosphere. *J. Meteor.*, **6**, 105—122.

Kuo, 1952: Three-dimensional disturbances in a baroclinic zonal current. *J. Meteor.*, **9**, 260—278.

— 1953: The stability properties and structure of disturbances in a baroclinic atmosphere. *J. Meteor.*, **10**, 235—243.

Rayleigh, Lord (J. W. Strutt), 1878: On the instability of jets. *Proc. Lond. Math. Soc.*, **10**, 4—13.

— 1880: On the stability or instability of certain fluid motions. *Proc. Lond. Math. Soc.*, **11**, 57—70.

— 1887: On the stability or instability of certain fluid motions, II. *Proc. Lond. Math. Soc.*, **19**, 67—74.

— 1913: On the stability of the laminar motion of an inviscid fluid. *Phil. Mag.*, **26**, 1001—1010.

Rossby, C.-G., 1939: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacement of the semi-permanent centers of action. *J. Mar. Res.*, **2**, 38—55.

Solberg, H., 1936: Le movement d'inertie de l'atmosphère stable et son rôle dans la théorie des cyclones. *Procès-verbaux*. L'assoc. météor., Un. géod. géophys. intern., Edimbourg (II), 66—82.

Sutcliffe, R. C., 1947: A contribution to the problem of development. *Quat. J. Roy. Meteor. Soc.*, **73**, 370—383.

Thompson, P. D., 1953: On the theory of large-scale disturbances in a two dimensional baroclinic equivalent of the atmosphere. *Quat. J. Roy. Meteor. Soc.*, **79**, 51—69.

Thomson, W. (Lord Kelvin), 1871: Hydrokinetic solutions and observations. *Phil. Mag.*, **42**, 362—377.

Univ. Chic. Dept. Meteor., 1947: On the general circulation of the atmosphere in middle latitudes. *Bull. Am. Meteor. Soc.*, **28**, 255—280.

The Spectrum of Large-scale Turbulent Transfer of Momentum and Heat

By MARIANO A. ESTOQUE[1, 2], Johns Hopkins University

(Manuscript received April 12, 1954)

Abstract

A method for obtaining the spectrum of turbulent transfer of a fluid property is presented based on cross-correlations and Fourier transforms. An application to large-scale eddy transfers of momentum and heat at 850 mb over the southeastern part of the United States shows that most of the transfers may be attributed to disturbances of periods from 0 to 10 days, the most important transfers occurring in the neighborhood of 4 days. Some implications of these results on the problem of the general circulation are discussed.

1. Introduction

One aspect of the general circulation which has attracted considerable interest in recent years concerns the nature of the processes which accomplish the required global balance of momentum and energy of the atmosphere. The necessity for the meridional transfer of angular momentum from tropical regions to middle latitudes was first pointed out by JEFFREYS (1926). He noted that since the prevailing winds at middle latitudes are westerlies, there is a constant abstraction of angular momentum from the atmosphere by frictional forces on the surface of the earth at these latitudes. On the other hand, the tropics, being regions of easterlies, are areas where angular momentum is supplied to the atmosphere. Since the zonal wind systems are quasi-stationary in the mean, there must be no progressive accumulation nor reduction of angular momentum at any one latitude belt. Hence, a continuous poleward flow of angular momentum from the source regions in the tropics to the sink at middle latitudes must exist. In a similar manner, if one considers the global heat balance, there is an excess of incoming solar radiation over outgoing radiation over equatorial regions; over the polar regions there is a net loss of heat. Therefore, in order to maintain the mean temperature difference between the poles and the equator relatively constant for long periods of time, there must exist a poleward transfer of heat.

For many years the mechanism whereby the atmosphere accomplishes these required transfers had been conceived along lines which depend upon convectively driven meridional circulation models originally proposed by HADLEY (1735). For a number of theoretical reasons, the validity of these models has been questioned. Consequently, emphasis is now being placed upon the theory that horizontal mixing processes are the mechanism which maintains the global balance of momentum and energy. Quantitative studies by WHITE (1951) and STARR and WHITE (1951) show that the

[1] This research has been made possible through support and sponsorship extended by the Geophysics Research Division of the Air Force Cambridge Research Center, under Contract No. AF 19 (122)—365.

[2] Now at the Dept. of Meteorology, University of Chicago.

meridional transfers of heat and momentum due to turbulent processes are capable of maintaining the required balance in middle latitudes. Thus, it is of prime importance in understanding the overall problem of the general circulation to have a quantitative knowledge of the distribution of the intensity of the turbulent transfers among the different scales of disturbances which comprise the entire spectrum of atmospheric turbulence. This paper is the result of investigations toward this direction. In the next section a method for determining the spectrum of turbulent transfer from observed data will be presented. The remaining sections give a numerical application and a discussion of some of the meteorological implications of the results.

Although the present study represents perhaps the first quantitative determination of the partition of turbulent flux among the different scales of eddies in the large-scale circulations of the atmosphere, a qualitative picture is readily obtained from synoptic considerations. An important study by DEFANT (1921) suggested that the traveling cyclones and anticyclones of middle latitudes might be considered as the individual "turbulence elements" in an intensive, quasi-horizontal process of exchange between air masses in high and low latitudes. In the work of Jeffreys cited, he also concluded that turbulence associated with these disturbances constitute the individual eddies which bring about a poleward flux of angular momentum. Along this line, STARR (1948) suggested that troughs in the westerlies which are oriented along a northeast-southwest direction produce the necessary velocity correlations to accomplish a net poleward flux of momentum. Aside from eddies of the scale of cyclones and anticyclones, there are fluctuations of much longer periods, such as those associated with variations in the zonal index, principally "blocking action", which may also produce net poleward transfers of momentum and energy. According to ELLIOT and SMITH (1949) the period of "blocking action" is from 15 to 30 days. While the role of "blocking action' in contributing to the meridional transport of momentum is difficult to visualize, studies by NAMIAS (1950) and ELLIOT and SMITH (op. cit.) emphasize their importance in the poleward transfer of heat. Based upon these considerations one would, therefore, anticipate the presence of maximum poleward eddy transfers in regions of the spectrum corresponding to a period of about 4 to 6 days which is approximately the mean interval between successive frontal or cyclone passages in middle latitudes. Another region of large transfers should be expected at longer periods to account for variations in the zonal index. Of course, the seasonal variations might also be of importance.

2. Theoretical Analysis for the Spectrum

The analysis which will be presented here will be developed through the use of cross-correlations and Fourier transforms. It must be mentioned that the results of such a development may be inferred from studies of the properties of general stationary random functions from the viewpoint of probability theory (BATCHELOR, 1953). However, the physical significance of a turbulent transfer spectrum is difficult to grasp solely with the aid of probability considerations. A more heuristic approach, although perhaps less rigorous, will be employed here by adopting a device introduced by WIENER (1930) of replacing an actual function by another function which differs from the former only in physically trivial aspects and extending ideas used by TAYLOR (1936) in obtaining the energy spectrum of velocity fluctuations.

Consider then a fluid element which has a velocity component V in the meridional direction and a property F. In the numerical application which will be given, F may either be a measure of heat or angular momentum such as absolute temperature, T, or the zonal velocity, U, respectively. Hence, the product VF is the meridional flux of the fluid property through an area normal to V. From a long and continuous record of observations of V and F over a station, let us choose two simultaneous intervals, one of each in V and F, having lengths, $2L$, sufficiently long to include a large number of fluctuations. The assumption regarding the length of the interval is necessary in order that one may obtain reasonably accurate information about statistical characteristics of the fluctuations. Let the origin of time be chosen at the midpoints of the intervals. Then it is evident from the manner in which the length of the intervals are determined that

the average values, $\overline{V}$ and $\overline{F}$, evaluated over the intervals must be very nearly equal to the corresponding averages over infinitely long periods of time, i.e.

$$\overline{V} = \frac{1}{2L}\int_{-L}^{L} V(t)dt \approx \lim_{A\to\infty}\frac{1}{2A}\int_{-A}^{A} V(t)dt \qquad (1)$$

$$\overline{F} = \frac{1}{2L}\int_{-L}^{L} F(t)dt \approx \lim_{A\to\infty}\frac{1}{2A}\int_{-A}^{A} F(t)dt \qquad (2)$$

Denoting the deviations from the means by the primed quantities

$$V'(t) = V(t) - \overline{V} \qquad (3)$$

$$F'(t) = F(t) - \overline{F} \qquad (4)$$

the mean turbulent or eddy transfer is

$$\overline{V'F'} = \frac{1}{2L}\int_{-L}^{L} V'(t)F'(t)dt = \overline{VF} - \overline{V}\,\overline{F} \qquad (5)$$

which shows that the mean transfer, $\overline{VF}$, may be considered as the sum of the mean turbulent transfer and a quantity, $\overline{V}\,\overline{F}$, which depends upon the existence of a mean meridional flow, $\overline{V}$. Recent calculations indicate that the contribution of $\overline{V}\,\overline{F}$ to the global momentum and heat balance is small compared to that of $\overline{V'F'}$, at least in the middle latitudes. An attempt will now be made to decompose the mean eddy transfer, $\overline{V'F'}$, into its spectral components.

For the purpose of analysis, let us replace the original set of fluctuations, $V'(t)$ and $F'(t)$, by a corresponding set denoted by $V^{\star}(t)$ and $F^{\star}(t)$ which would still retain the general statistical characteristics of the original fluctuations. Such a set will be defined for all values of time by the equations

$$V^{\star}(t) = \begin{cases} V'(t) & -L \le t \le L \\ 0 & t \ge |L| \end{cases} \qquad (6)$$

$$F^{\star}(t) = \begin{cases} F'(t) & L \le t \le L \\ 0 & t \ge |L| \end{cases} \qquad (7)$$

The implication of this artifice is that the original fluctuations, which may persist for an infinite period of time, are replaced by the finite segments of the original fluctuations which occur over the intervals, $-L \le t \le L$, and flanked on both sides by infinitely long intervals with fluctuations which are identically zero. It may be shown that the primary effect of this mathematical expediency is a tendency to "smear" the spectrum. For example, an actual line at $\omega = \omega_0$ in the spectrum would appear like an error curve, or more precisely, the function $\dfrac{\sin(\omega - \omega_0)}{\omega - \omega_0}$ in the analysis. The "resolving power" of the analysis is directly proportional to the length, $2L$, of the interval.

Let us express the fluctuations given by equations (6) and (7) as the Fourier integrals,

$$V^{\star}(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g_v(\omega)e^{i\omega t}d\omega \qquad (8)$$

where

$$g_v(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} V^{\star}(t)e^{-i\omega t}dt \qquad (9)$$

and

$$F^{\star}(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g_F(\omega)e^{i\omega t}d\omega \qquad (10)$$

$$g_F(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} F^{\star}(t)e^{-i\omega t}dt \qquad (11)$$

Using Parseval's theorem for Fourier integrals, one obtains

$$\int_{-\infty}^{\infty} V^{\star}(t)\hat{F}^{\star}(t)dt = \int_{-\infty}^{\infty} g_v(\omega)\hat{g}(\omega)d\omega \qquad (12)$$

the circumflex indicating that the quantities are complex conjugates. Noting that $V^{\star}(t)$ and $F^{\star}(t)$ are real

$$\int_{-\infty}^{\infty} V^{\star}(t)\hat{F}^{\star}(t)dt = \int_{-\infty}^{\infty} V^{\star}(t)F^{\star}(t)dt = \int_{-L}^{L} V'(t)F'(t)dt \approx 2L\overline{V'F'} \qquad (13)$$

where $\overline{V'F'}$ is the average value of the turbulent transfer taken over an infinitely long record of the original set of fluctuations. Therefore, with the aid of equation (12)

$$\overline{V'F'} = \frac{1}{2L}\int_{-\infty}^{\infty} g_v(\omega)\hat{g}_F(\omega)d\omega = \int_{-\infty}^{\infty} Q(\omega)d\omega \qquad (14)$$

$$Q(\omega) \equiv \frac{g_v(\omega)\hat{g}_F(\omega)}{2L} \qquad (15)$$

which shows that the turbulent transfer may be expressed as contributions of a continuous spectrum where ω is the frequency. The quantity, $Q(\omega)$, contains a measure of the intensity of eddy transfer per unit bandwidth of frequency.

In order to express $Q(\omega)$ in terms of the original fluctuations, one may proceed in the following manner. Using the relation between the Fourier transform of the product of two functions and the "Faltung" of their Fourier transforms,

$$\int_{-\infty}^{\infty} g_v(\omega)\hat{g}_F(\omega)e^{i\omega\tau}d\omega = \int_{-\infty}^{\infty} V^{\star}(t)F^{\star}(t+\tau)dt$$

$$= \int_{-L}^{L} V'(t)F'(t+\tau)dt \qquad (16)$$

and with the aid of the definition for $Q(\omega)$,

$$\int_{-\infty}^{\infty} Q(\omega)e^{i\omega\tau}d\omega = \frac{1}{2L}\int_{-L}^{L} V'(t)F'(t+\tau)dt =$$

$$= \overline{V'(t)F'(t+\tau)} \qquad (17)$$

It is worthy to note again that, since the interval, $2L$, contains a large number of the fluctuations, the mean over $2L$, $\overline{V'(t)\,F'(t+\tau)}$, must be an excellent approximation of the actual value of the cross-correlation with lag τ evaluated for an infinitely long record of the original fluctuations. The preceding equation resembles one member of a set of integrals defining a Fourier integral and its transform. Therefore, by analogy

$$Q(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{V'(t)F'(t+\tau)}e^{-i\omega\tau}d\tau \qquad (18)$$

Hence, the spectrum of turbulent transfer can be obtained as a Fourier transform of the cross-correlation function. In general, $Q(\omega)$ is a complex quantity. Its real part, $Q_r(\omega)$, corresponds to the "amplitude spectrum" and represents one-half of what will be defined later as the spectrum of turbulent transfer. The imaginary part, $Q_i(\omega)$, which is physically less significant in our particular study, represents the "phase spectrum". Since the cross-correlation is real, the real part of the spectrum is

$$Q_r(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{V'(t)F'(t+\tau)}\cos\omega\tau d\tau \qquad (19)$$

By following exactly the same procedure used in the preceding analysis except for interchanging the role of $V'(t)$ and $F'(t)$, one can also show that

$$Q_r(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{V'(t)F'(t-\tau)}\cos\omega\tau d\tau \qquad (20)$$

Adding these two expressions for $Q_r(\omega)$ one derives

$$Q_r(\omega) =$$

$$= \frac{1}{4\pi}\int_{-\infty}^{\infty} \left[\overline{V'(t)F'(t-\tau)} + \overline{V'(t)F'(t+\tau)}\right]\cos\omega\tau d\tau \qquad (21)$$

Since the integrand is now an even function of τ

$$Q_r(\omega) =$$

$$= \frac{1}{2\pi}\int_{0}^{\infty} \left[\overline{V'(t)F'(t-\tau)} + \overline{V'(t)F'(t+\tau)}\right]\cos\omega\tau d\tau \qquad (22)$$

Note that $Q_r(\omega)$ is also an even function, i.e., $Q_r(\omega) = Q_r(-\omega)$. The sum of $Q_r(\omega)$ and $Q_r(-\omega)$ representing the total contribution due to any frequency, ω, regardless of sign, will be termed the spectrum of turbulent transfer, $\Phi(\omega)$. Thus,

$$\Phi(\omega) =$$

$$= \frac{1}{\pi}\int_{0}^{\infty} \left[\overline{V'(t)F'(t+\tau)} + \overline{V'(t)F'(t-\tau)}\right]\cos\omega\tau d\tau \qquad (23)$$

One may again write

$$\left.\begin{aligned}\overline{V'(t)F'(t+\tau)} + \overline{V'(t)F'(t-\tau)} = \\ = 2\int_0^\infty \Phi(\omega)\cos\omega\tau d\omega\end{aligned}\right\} \quad (24)$$

by observing that the resulting forms of the equations are analogous to a Fourier transform and its inverse. In particular, if τ is zero, the preceding equation reduces to an expression for the mean eddy transfer in terms of its spectrum,

$$\overline{V'F'} = \int_0^\infty \Phi(\omega)d\omega \quad (25)$$

showing that the total area under the spectral curve is equal to the mean turbulent transfer.

In practical applications, it is sometimes more convenient to evaluate directly an integral of the spectral curve, representing the total transfer due to particular frequency bands, and then derive the actual spectrum by differentiating the integrated spectrum. For this purpose, the contribution of all eddies whose frequencies range from ω_1 to ω_2 may be represented by the expression

$$S(\omega_1 \to \omega_2) = \int_{\omega_1}^{\omega_2} \Phi(\omega)d\omega \quad (26)$$

or substituting the expression for $\Phi(\omega)$ in (23)

$$S(\omega_1 \to \omega_2) = \\ = \frac{1}{\pi}\int_0^\infty \left[\overline{V'(t)F'(t+\tau)} + \overline{V'(t)F'(t-\tau)}\right] \cdot \\ \cdot \frac{\sin\omega_2\tau - \sin\omega_1\tau}{\tau} d\tau \quad (27)$$

3. The Integrated Spectrum of Turbulent Transfer for Heat and Momentum.

An application of the theoretical analysis in the preceding section was made based on radiosonde soundings obtained twice daily over Greensboro (36° N, 80° W) in the southeastern part of the United States for the year 1949. Since the numerical work was fairly extensive it was possible to treat only the data at the 850 mb level. It was also convenient to use geostrophic winds instead of the actual observed winds because the former were readily available in forms suitable for computation. In order to avoid unnecessary numerical work, constants involving density, specific heat at constant pressure, and the distance of the station from the earth's axis are omitted. Heat and angular momentum flux are denoted simply by VT and UV, respectively. To convert them into the proper units the appropriate physical constants for the set of observations must be applied.

At the start of the computations it was believed that the annual eddy would contribute sizable transfers. Consequently, it was decided to minimize the possible "smearing effect" of the annual cycles in U, V, and T by estimating their components using harmonic analysis. These components, together with the corresponding annual means, were subsequently subtracted from the original observations to obtain the fluctuations, U', V', and T'. The average cross-correlations,

$$\frac{\overline{V'(t)T'(t+\tau)} + \overline{V'(t)T'(t-\tau)}}{2}$$

and

$$\frac{\overline{V'(t)U'(t+\tau)} + \overline{V'(t)U'(t-\tau)}}{2}$$

were then computed from the fluctuations for convenient lags. The results up to a lag of 100 days are shown in Fig. 1. It can be seen that both curves are similar for small values of lags. The correlations decrease very rapidly from their values at zero lag indicating that the most effective eddies in transporting heat and momentum are short period disturbances. At larger lag periods the curves have no resemblance to each other. It had been expected that the correlations would become extremely small at long lag periods of the order of 100 days. Actually this was not realized on account of sampling errors introduced by the relatively smaller number of cases from which the means were evaluated at these lags. Although the correlations do become quite small near lags of 100 days, beyond this, the correlations become more and more unstable. This is a reflection of the fact that, for a given set of observations, the number of cases used in computing a cross-correlation decreases in direct proportion to the length of the lag period. It is, therefore, quite possible that for the longer lag periods,

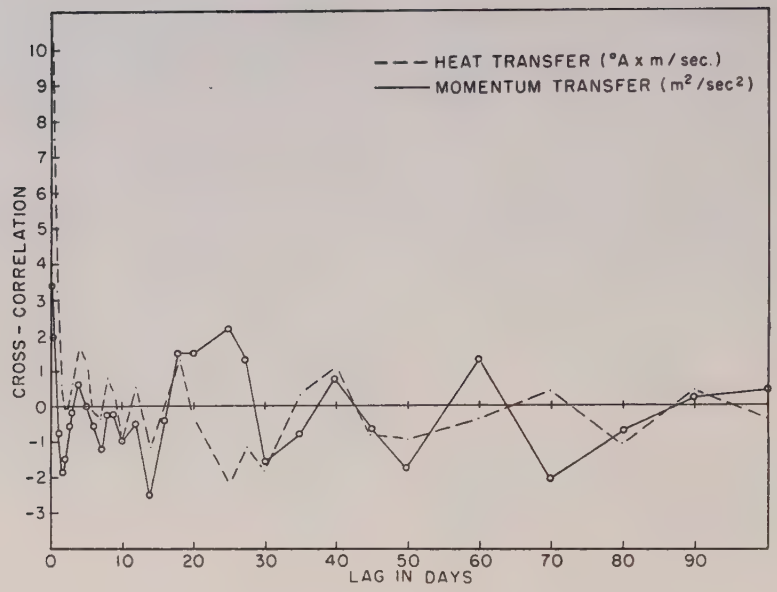


Fig. 1. The average cross-correlations for heat and momentum.

the correlations are highly unrepresentative; this will result in an inaccurate determination of the spectrum at longer periods. In view of these observations on the uncertainties of these particular correlations, it was decided to minimize their effects as much as possible by evaluating the integrated spectrum representing the total contribution of all eddies having frequencies from 0 to ω, instead of the actual spectrum. Thus, using equation (27) and expressing the frequency in terms of the period, $p = \frac{2\pi}{\omega}$

$$S(0 \to \omega) = \frac{2}{p}\int_0^{\infty} \left[\overline{V'(t)F'(t+\tau)} + \overline{V'(t)F'(t-\tau)}\right] \cdot \frac{\sin\frac{2\pi\tau}{p}}{\frac{2\pi\tau}{p}}\, d\tau \qquad (28)$$

For large values of τ the factor $\sin\frac{2\pi\tau}{p} \Big/ \frac{2\pi\tau}{p}$ becomes small, having a value of at most $\frac{p}{2\pi\tau}$; the whole integrand becomes small and its contribution to the value of the integral is negligible. Therefore, the unreliability of the cross-correlations at large lags are minimized.

Figures 2 and 3 represent the integrated spectra for the turbulent transfers of heat and

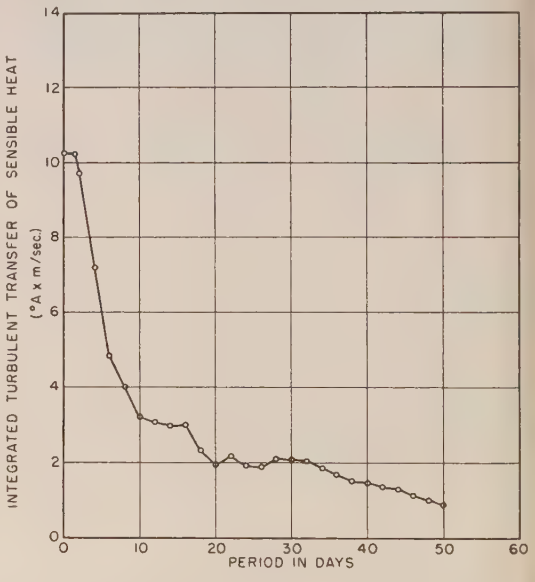


Fig. 2. The integrated spectrum for turbulent heat transfer. The ordinate represents the total transfer due to all disturbances with periods longer that than given by the corresponding abscissa.

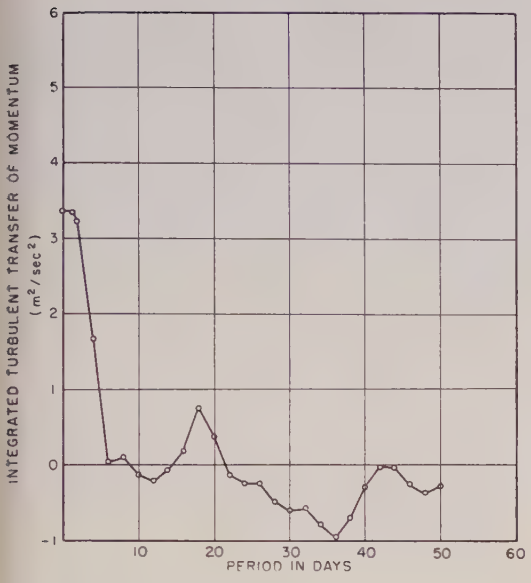


Fig. 3. The integrated spectrum for turbulent momentum transfer. The ordinate represents the total transfer due to all disturbances with periods longer than that given by the corresponding abscissa.

momentum obtained with the aid of equation (28) after replacing the integration by a summation. In order to show more clearly the relative importance of the different scales of disturbances in the turbulent transfer of heat and momentum, the integrated spectra were smoothed and then graphically differentiated. The results given in Figure 4 show the intensity of transfer as a function of period.

The most striking characteristic of both spectra is the extreme importance of short-period disturbances in the eddy transfer of heat and momentum. The high peaks at about 4 days are undoubtedly the expected contributions of frontal disturbances and the associated anticyclones. The total amount of transfers due to eddies with periods larger than 10 days is relatively small in the case of heat flux; and for momentum transfer the net contribution of these long-period eddies is entirely negligible for this particular set of data due to the presence of approximately equal amounts of northward and southward transfers. This effect is extremely slight for heat transport and one might infer that the tendency of all kinds of atmospheric disturbances, irrespective of size, is to effect poleward heat transfers. The differences in the characteristics of the two spectra might be a result of the different effectiveness of zonal index variations with regard to the ability to transfer heat and momentum. It must be emphasized again that the results of the spectrum analysis for longer periods are not entirely reliable so that these should be interpreted with caution.

The integrated spectra show that the total turbulent transfer due to all eddies excluding the yearly eddy are 3.35 $m^2\ sec^{-2}$ and 10.21 $A\,x\,m\ sec^{-1}$ for momentum and heat, respectively. Since the corresponding total transfers including the yearly eddy as evaluated from the original data are 3.54 $m^2\ sec^{-2}$ and 10.80 $A\,x\,m\ sec^{-1}$, the contributions of the yearly eddy are approximately 3 per cent and 6 per cent for momentum and heat, respectively. This comparatively small contribution of the yearly eddy in the meridional flux of these quantities is also true for the turbulent transfer of latent heat as shown by recent studies (Benton et al., 1953). Consequently, the evaluation of the local eddy flux with the aid of averaging methods (using an equation of the form, $\overline{V'F'} = \overline{VF} - \bar{V}\bar{F}$) for periods of at least one month would include the transfers of most of the disturbances of the entire spectrum. The dominant role of short-period disturbances is more clearly illustrated in Table I which shows the ratio of the transfers due

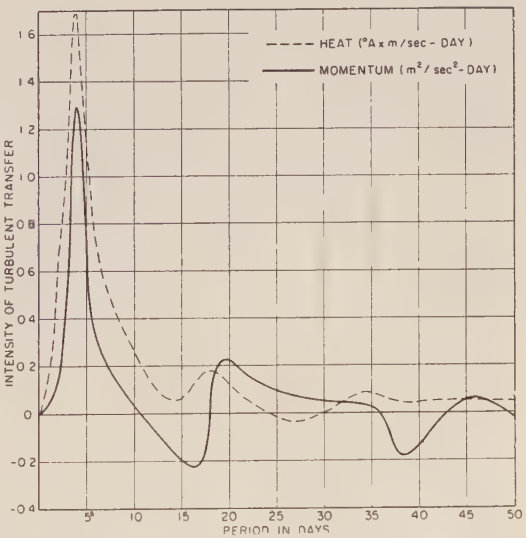


Fig. 4. The intensity of heat and momentum transfers plotted against period.

to different bands of the spectrum to the total turbulent transfer.

Table 1. Turbulent transfers due to different regions of the spectrum expressed as percentage of the total transfer. Negative percentages indicate southward transfers.

Period (days)	0—10	10—20	20—30	30—40	40—50	over 50
Momentum (per cent)	98	—14	28	—9	0	— 3
Heat (per cent) ..	65	12	—1	6	5	13

The prominent regions in the momentum spectrum showing appreciable transfers at longer periods are quite interesting. These characteristics of the spectrum are difficult to associate definitely with any large-scale disturbances in the atmosphere and may possibly be peculiar to the particular set of data analyzed or due to the unreliability of results at the longer periods. A more complete analysis of other cases should indicate whether these peculiarities in the spectrum are real and local in character.

4. Summary and Conclusions

Perhaps an important result of the present study is the fact that the method of spectrum analysis based on correlations is practicable and useful in deducing the nature of large-scale turbulent transfer processes such as those involved in the general circulation. At this point it is worthwhile to mention some points which will serve as guides in future applications of the method. Our numerical work revealed that the most critical aspect of the computations involved the determination of representative cross-correlations. It was found that there is difficulty in obtaining reliable values of the correlations for lags of the order of 100 days by using only one year of data consisting of two observations per day. The only solution to this difficulty is to make either more frequent observations or use longer periods of observation. The former is preferred because the analysis would then enable one to assess readily the transporting ability of disturbances with periods of fractions of a day. In order to minimize "smearing", the removal of the components of the annual cycle from the original set of observations is necessary if its contribution is not negligible as compared to the total turbulent transfer. A rough idea of the transfer due to the annual cycle is given by the formula $\frac{AB}{2}\cos\Theta$, where A and B are the amplitudes of the annual waves in V and F obtained from monthly means and Θ is the difference of their phase angles. Finally, it is not advisable to compute the spectrum directly using equation (23) unless the reliability of the cross-correlations is completely assured. A better procedure is to evaluate the integrated spectrum first and to derive the spectrum by differentiation.

It is certainly not appropriate to make definite conclusions about the nature of horizontal mixing processes which accomplish the balance of momentum and heat required by the general circulation problem from the very limited application of the spectrum analysis presented above. Our work must be extended first to include observations at upper levels of the troposphere as well as at other latitudes. A spectrum analysis at 300 mb over the same station considered here is particularly desirable since it is the level of maximum northward transport of angular momentum and hence may indicate the nature of the principal mechanism of momentum balance. It is also extremely interesting to apply the analysis over tropical latitudes where the atmospheric processes accomplishing poleward eddy transfers of heat and momentum are least understood. Recent work by STARR and WHITE (1952) indicates that the local eddy flux over these regions are not at all negligible as it was presumed to be.

In spite of the limited scope of our numerical application, however, it is possible to offer some comments regarding the nature of atmospheric processes which accomplish the required poleward transfers of momentum and heat energy. In the lower portions of the troposphere, and possibly including the 500 mb level, the dominant mechanisms are those associated with frontal disturbances and migratory cyclones and anticyclones. This fact lends support to the theory that frontal disturbances and the associated cyclones and anticyclones are indispensable processes in the maintenance of the general circulation. For the local eddy flux of enthalpy and latent heat, these disturbances are the most effective throughout

the entire vertical extent of the atmosphere at middle latitudes since heat and moisture flux above 500 mb is relatively weak. The same statement cannot, of course, be made in the case of momentum flux. The level of maximum transfers occurs at about 300 mb, much higher than the level treated in the present study. It is possible that at this level the region of most intense transfers, which were found near periods of 4 days at 850 mb, may be shifted to longer periods as a result of the effect of long waves in the upper westerlies.

Acknowledgements

— The author wishes to express his gratitude to Dr George S. Benton and the members of the Water Vapor Transport Project of the Johns Hopkins University for their invaluable help extended during the investigation.

REFERENCES

Batchelor, G. K., 1953: *The Theory of Homogeneous Turbulence*. Cambridge Univ. Press, Cambridge, 197 pp.

Benton, G. S., Estoque, M. A., and Dominitz, J., 1953: An evaluation of the water vapor balance of the North American continent. *Scientific Report No. 1*, Department of Civil Engineering, the Johns Hopkins Univ., Baltimore, Maryland.

Defant, A., 1921: Die Zirkulation der Atmosphäre in den gemässigten Breiten der Erde. *Geogr. Ann.*, **3**, 209—266.

Elliot, R. C., and Smith, T. B., 1949: A study of the effects of large blocking highs on the general circulation in the northern hemisphere westerlies. *J. Meteor.*, **6**, 67—85.

Hadley, G., 1735: Concerning the cause of the general trade winds. *Phil. Trans. Roy. Soc. London*, 39: 58. Reprinted in "The mechanics of the earth's atmosphere. A collection of translations by Cleveland Abbe", 3rd Collection. Smithsonian Misc. Coll., 51: 4, 5—7.

Jeffreys, H., 1926: On the dynamics of geostrophic winds. *Quart. J. R. Meteor. Soc.*, **52**, 85—104.

Namias, J., 1950: The index cycle and its role in the general circulation, *J. Meteor.*, **7**, 130—139.

Starr, V. P., 1948: An essay on the general circulation of the earth's atmosphere. *J. Meteor.*, **5**, 39—43.

— and White, R. M., 1951: A hemispherical study of the atmospheric angular momentum balance. *Quart. J. R. Meteor. Soc.*, **7**, 215—225.

—, —, 1952: Meridional flux of angular momentum in the tropics. *Tellus*, **4**, 118—125.

Taylor, G. I., 1936: The spectrum of turbulence. *Proc. Roy. Soc. London*, Ser. A, **164**, 476—490.

White, R. M., 1951: The meridional eddy flux of energy. *Quart. J. R. Meteor. Soc.*, **77**, 188—199.

Wiener, H., 1930: Generalized Harmonic Analysis. *Acta Math.*, **55**, 117—258.

Note on the Release of Kinetic Energy in Tropical Cyclones

By E. PALMÉN and C. L. JORDAN, University of Chicago[1]

(Manuscript received September 27, 1954)

The rate of change of horizontal kinetic energy in a given volume can be expressed by

$$\frac{\partial}{\partial t}\int_V K dV = -\int_S K v_r dS -$$

$$-\int_V g\varrho\left(u\frac{\partial H}{\partial x}+v\frac{\partial H}{\partial y}\right)dV-\int_V F dV, \quad (1)$$

where K denotes the horizontal kinetic energy per unit volume, v_r is the outward component of velocity at the boundary, g is the acceleration of gravity, ϱ is the density, u and v denote the horizontal components of velocity and H is the height of a given isobaric surface. In Eq. (1) dV further denotes an element of volume and dS an element of the boundary surface. The third term on the right represents the rate at which all kinds of friction are decreasing the kinetic energy per unit time. Since the contribution of the vertical kinetic energy to the total kinetic energy is in general small, K can be considered as the total kinetic energy per unit volume.

Eq. (1) states that the rate of change of kinetic energy in a given volume V is determined by the flux of kinetic energy across the boundary, the generation of kinetic energy within the volume itself and the dissipation of kinetic energy in the volume due to frictional forces. In a steady state the latter must be equal to the generation reduced by the outflow of kinetic energy.

A tropical cyclone in the mature stage represents closely a steady state in a coordinate system moving at the mean speed of the cyclone. Considering a cyclone with symmetrical distribution of wind and pressure about a vertical axis in the center of the eye, the second term on the right of (1) can be written:

$$-A\int_0^{p_0}\overline{v_r\frac{\partial H}{\partial r}}\,dp. \quad (2)$$

Here p_0 denotes the pressure at the surface, v_r is the velocity in the direction of the radius r, and A is the area limited by the circle considered. The bar indicates a "mean value" of the product $v_r\frac{\partial H}{\partial r}$ in the total area A. Expression (2) represents the generation of kinetic energy per unit time in a cylinder with the horizontal area A extending from the earth's surface to the upper limit of the atmosphere. As the integration is extended over larger volumes the contribution from the first term on the right in Eq. (1) decreases rapidly and over a sufficiently large volume probably becomes negligible. Then the total dissipation of energy can be considered approximately equal to the generation computed from (2).

The generation of kinetic energy in a cyclone of the symmetric type discussed here can be evaluated from the integral (2). Since v_r is negative at lower levels and positive at higher levels the condition for generation of kinetic energy is that the slope of the isobaric surfaces in the cyclone must, on the average, be steeper in the layers of inflow than in the layers of outflow. This is possible only if the interior of the cyclone with ascending motion is warmer than the outer region. Such a distribution of temperature is possible as long as the surface temperature remains sufficiently high, as has been shown by PALMÉN (1948).

[1] Prepared under a research contract between the Office of Naval Research and the University of Chicago.

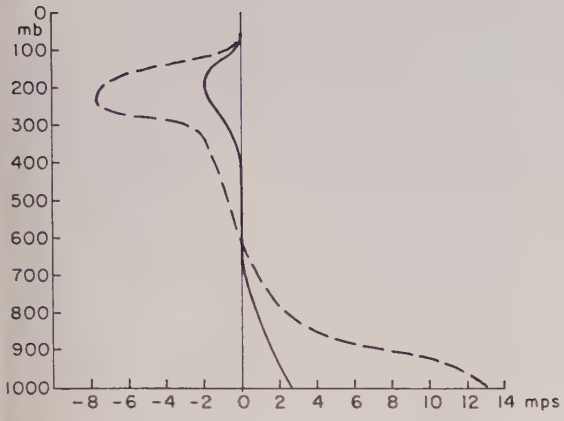


Fig. 1. Vertical distribution of the negative radial velocity component in the "mean tropical cyclone" for the 0.25—2.0° lat. ring (dashed curve) and the 5.0—6.0° lat. ring (solid curve).

In the following an attempt will be made to compute the integral (2). Since the data are not adequate for any individual case, the application will be made to a "mean tropical cyclone" as defined by the mean wind and pressure-height data presented by HUGHES (1952) and E. S. and C. L. JORDAN (1952, 1954). The use of mean data raises some difficulties, but since our computation only intends to give an idea of the order of magnitude of the mean release of kinetic energy, these difficulties should not be critical. In fact, since individual tropical cyclones over the warm ocean areas are quite similar, a storm of the size defined by the above investigations could be expected to show wind and pressure distributions close to those given for the mean case.

The distribution of v_r and $\frac{\partial H}{\partial r}$ given by the basic data of the "mean tropical storm" was not complete enough in all respects for the evaluation of (2). Certain extrapolations of these data were necessary; for example, the winds at the lowest level—1,000 ft—extended to a distance of 0.5° lat. from the storm center, but at upper levels only to a distance of 2° lat. Likewise, no data were available for the uppermost 100 mb of the atmosphere. The mean wind data were fitted together in a consistent manner by imposing the mass continuity requirement for each concentric ring. Greater reliance was placed in the strong inflow near the surface and the strong outflow in the upper troposphere since the individual data in these areas showed good consistency. Data for the layer 600—300 mb were not considered very reliable since the individual wind reports here ranged from strong inflow to strong outflow.

Mean values of v_r and $\frac{\partial H}{\partial r}$ were determined at each level[1] for the rings with the following radii: 0.25—2° lat.,[2] 2—3°, 3—4°, 4—5°, and 5—6°. These values were plotted against a linear pressure scale and smooth curves were drawn. The curves of v_r and the product $v_r \frac{\partial H}{\partial r}$ for the inner and outer rings are shown in Figs. 1 and 2. From the latter figure the production of kinetic energy per unit area in the different layers can be computed graphically. The total production for each ring can then be determined by summing the individual layers and multiplying by the appropriate area.

In the inner zone the layer of strong inflow coincides with the regions of highest values of $\frac{\partial H}{\partial r}$, hence the major contribution to the kinetic energy increase appears here. At upper levels the outflow is toward higher

[1] Wind data were available at 1000, 4000, 7000, 10000, 18000, 30000, 40000 and 45000 ft and pressure-height data at 1000, 850, 700, 500, 400, 300, 200 and 100 mb.

[2] The eye was assumed to have a diameter of 0.5° lat.

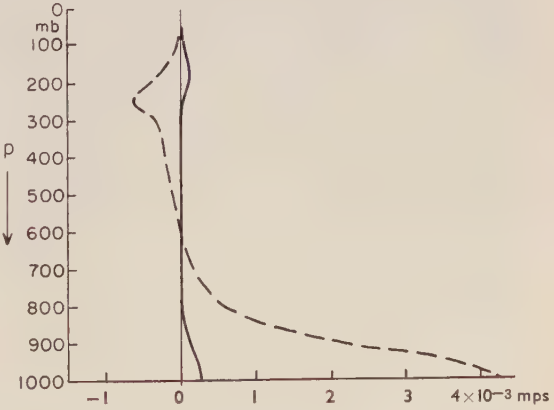


Fig. 2. Vertical distribution of the product $-\overline{v_r \frac{\partial H}{\partial r}}$ in the "mean tropical cyclone" for the 0.25—2.0° lat. ring (dashed curve) and 5.0—6.0° lat. ring (solid curve).

Table I. Production of kinetic energy (unit: 10^{24} ergs/day) in different rings of a mean tropical cyclone compared to the release of latent heat in the interior of the storm

Ring	Release of kinetic energy	Release of latent heat
0.25—2.0° lat.	4.1	419
2.0 —3.0	2.5	17
3.0 —4.0	2.5	29
4.0 —5.0	2.2	—
5.0 —6.0	1.6	—
0.25—6.0	12.9	about 500

pressure near the storm center and therefore energy consuming, but in the outer regions is toward lower pressure and thus contributes to an increase in the kinetic energy. However, even after considering the greater area in the outer ring this contribution is small in comparison to the production of kinetic energy in the regions of strong inflow near the surface in the interior of the cyclone.

The net increase of kinetic energy in ergs per day for the various rings, when frictional dissipation and flux of energy through the boundaries are neglected, is shown in Table I. For comparison the release of latent heat, as computed by HUGHES (1952) is presented in the last column. This release of latent heat is almost entirely limited to the inner region of the cyclone (the rain area). A rough calculation of the radial transport of kinetic energy at the 6° lat. ring showed essentially no net transfer. Thus the accumulative total generation of kinetic energy inside the radius of 6° lat., roughly 13×10^{24} ergs/day, also approximates closely the frictional dissipation in the same volume, since in our mean cyclone the energy can be considered constant. If the total release of latent heat inside the ring at 6° lat. is roughly estimated to be 500×10^{24} ergs/day by extrapolating the computations made by Hughes, the generation of kinetic energy in a typical tropical cyclone can be estimated as about 1/40 the release of latent heat energy due to ascent of moist air near the storm center. This result is in agreement with estimates made by RIEHL (1954) and also with computations of the dissipation of kinetic energy due to friction in the surface layer (HORIGUTI, 1928; HUGHES, 1952). Since these authors did not consider the total frictional dissipation it seems likely that they have somewhat overestimated the surface friction. However, considering the vertical distribution of wind velocity it is probable that a larger proportion of the frictional dissipation occurs in the surface layer in tropical cyclones than in extratropical storms where the wind maximum is usually near the tropopause level.

The total kinetic energy in the "mean tropical cyclone" was computed by determining the areal mean value of $v_s^2 + v_r^2$ at each level and performing a graphical integration from the surface to the upper limit of the atmosphere. The computed kinetic energy was 9×10^{24} ergs inside the 4° ring and 12×10^{24} ergs inside the 6° ring. The latter value is roughly equal to the daily generation of kinetic energy in the mean cyclone according to Table I. Therefore the half-life of the tropical cyclone—the time the cyclone would require to lose half its kinetic energy if its energy source were to disappear—is somewhat less than one day. Although the use of mean values in such a computation must in general lead to a kinetic energy total which is too small, the half-life obtained is certainly reasonable considering the rapid dissipation of intense tropical cyclones which occurs with the movement of colder air into the storm circulation at lower levels (LA SEUR and JORDAN, 1952). This result seems also to be in agreement with the ideas recently expressed by BERGERON (1954) concerning the filling of tropical storms moving in over land.

REFERENCES

BERGERON, T., 1954: The problem of tropical hurricanes. *Quart. J. R. Meteor. Soc.*, **80**, 131—164.

HORIGUTI, Y., 1928: On the typhoon of the Far East. *Mem. imp. Observ., Kobe*, **3**, 23—65.

HUGHES, L. A., 1952: On the low-level wind structure of tropical storms. *J. Meteor.*, **9**, 422—428.

JORDAN, E. S., 1952: An observation study of the upper wind-circulation around tropical storms. *J. Meteor.*, **9**, 340—346.

JORDAN, C. L., and JORDAN, E. S., 1954: On the mean thermal structure of tropical cyclones (To be published in the *Journal of Meteorology*).

LA SEUR, N. E., and JORDAN, C. L., 1952: A typical weather situation of the typhoon season. *Dept. Meteor., Univ. Chicago*, 24 pp.

PALMÉN, E., 1948: On the formation and structure of tropical hurricanes. *Geophysica*, **3**, 26—38.

RIEHL, H., 1954: *Tropical Meteorology*, New York, McGraw-Hill Book Co., 392 pp.

A Two-parameter Representation of the Normal Temperature Distribution of the 1,000—500 mb Layer

By G. ÁRNASON[1] and L. VUORELA[2], University of Stockholm

(Manuscript received June 14, 1954)

Abstract

Height normals for 1,000, 700 and 500 mb, published by U.S. Weather Bureau, have been used to study the normal temperature field of the 1,000—500 mb layer by means of a simple model. In this model the normal temperature-height curve is replaced by one having a temperature lapse-rate independent of height in such a way that the thicknesses of the sub-layers 1,000—700 and 700—500 mb are correctly represented by the model. As a consequence, the temperature as well as its horizontal gradient and the thermal wind all have a linear variation with height and may be described in terms of two parameters, the temperature lapse-rate k and a "representative" 1,000 mb temperature τ_0.

Hemispheric maps for these parameters as well as the thermal wind at 1,000 and 500 mb have been prepared for the months January and July. The major frontal zones of the atmosphere appear clearly on the τ_0-maps in positions which agree essentially with those shown earlier by a number of meteorologists. The July map reveals, however, a strong baroclinic zone in high latitudes, surrounding the polar cap, not shown clearly in earlier presentations of the low-level baroclinity.

The k-charts disclose large horizontal variation in the temperature lapse-rate, and the charts for the thermal wind show that its variation with height is very pronounced in middle and high latitudes.

1. Introduction

Until recently, reliable hemispheric charts displaying the normal distribution of various meteorological elements have not been available, except for surface data and those meteorological phenomena which could be observed from the ground. The greatly improved network of upper air stations since World War II has made it possible to carry out regularly contour and temperature analysis of selected pressure surfaces on a hemispheric scale, and hence to prepare upper air maps which may be referred to as normals. It is obviously of great importance to obtain as soon as possible a fairly complete picture of the different meteorological elements. Once a three-dimensional picture of temperature and pressure has been arrived at, the distribution of such derived quantities as temperature lapse-rate, geostrophic wind, geostrophic vorticity, thermal wind and thermal vorticity is easily obtained. More advanced aspects are also visualized; high-speed electronic computers make it possible to compute the three-dimensional fields of the vertical velocity and horizontal divergence associated with the current weather maps, and hence to prepare normal maps for these quantities. If one is justified in extrapolating the present trend in numerical forecasting, it seems likely that detailed upper air information will have to be made available on a routine basis for the purpose of numerical prediction and

[1] Present address: Joint Numerical Prediction Unit, Washington D. C.

[2] Present address: University of Helsinki, Institute of Meteorology, Helsinki.

in a form suitable for the computations of the fields of vertical velocity and horizontal divergence. Although the distribution of these quantities, as expressed by the present atmospheric models used for the purpose of numerical forecasting, may be too crude for the preparation of normals, it is not difficult to design a more refined model for this particular purpose. The efforts needed in order to produce the additional information required by such a model may not be prohibitive from a practical point of view. The availability of current maps for vertical velocities would seem to be an urgent requirement for a complete and successful study of the general circulation.

A question which naturally arises in the discussion of three-dimensional normals, is the form in which these normals should be presented, and in particular whether the usual form is the most suitable one. The present trend would seem to be to give the normal distribution of a certain meteorological element in as many as possible of the 7 pressure levels (1,000, 850, 700 mb etc.) currently reported in routine observations; this, of course, constitutes a fairly complete three-dimensional picture. Still the main characteristics of the vertical distribution may be more easily examined if described by a few properly selected parameters than in the 7-chart presentation. It seems, for instance, likely that the vertical velocity distribution along the vertical could be fairly accurately described in terms of 3 or 4 parameters. A more obvious example is the vertical distribution of temperature with height. The approximate linearity in the normal temperature-height curve suggests immediately the most suitable choice of parameters.

One of the co-authors has shown elsewhere (ÁRNASON, 1952 and 1953) that the main characteristics of the vertical temperature distribution may be described in terms of three parameters, namely, the average temperature lapse-rate in the troposphere, a "representative" sea-level temperature, and the height of the tropopause. A description of this kind gives a clearer picture of the normal distribution of temperature than if displayed by means of the 7 above-mentioned charts.

A treatment of this kind has been applied to height and thickness normals published by the U.S. WEATHER BUREAU (1952) with the aim of studying the main features of the three-dimensional distribution of temperature within the layer 1,000—500 mb. Two months, January and July, were selected, representing a typical winter and summer conditions respectively. The normal temperature-height curve was approximated by a straight line which was adjusted so that the 1,000—700 and 700—500 mb thicknesses were correctly given by the model. Two parameters are needed for the description of the temperature field, the lapse-rate k and the "representative" 1,000-mb temperature τ_0.

Chapter 2 gives the details pertaining to the determination of the two parameters; chapters 3 and 4 are devoted to the discussion of the k- and τ_0-maps, whereas the corresponding thermal wind field is discussed in chapter 5. The main achievement of the investigation laid down in the subsequent chapters is in short as follows:

1. It shows the usefulness of applying a simple model to meteorological normals, presented in the conventional manner, in order to obtain a clear picture of the main characteristics of the vertical distribution.
2. It describes in considerable detail the normal fields of the temperature lapse-rate and the thermal wind, neither of which has been studied earlier to the same extent.
3. It gives a measure of the baroclinity of the lowest layer of the atmosphere which is more representative than the observed temperature of the earth's surface, and which more clearly than the 1,000—700 mb thickness isopleths shows the major frontal zones of the atmosphere.
4. The results strongly suggest that the variation in the thermal wind with height is more pronounced than commonly believed, and that this feature should be given due consideration in the design of models for numerical forecasting.

2. Method of determining the temperature parameters

In the model, referred to in the introduction, the observed normal distribution of temperature with height in the troposphere is replaced by the linear relationship

$$\tau = \tau_0 - kz \qquad (2.1)$$

where τ is the temperature at height z, k the temperature lapse-rate (independent of height), and τ_0 the temperature at $z=o$. The temperature τ_0 may be referred to as a "representative" sea level temperature since it is, as will appear below, much less influenced by the local heat and cold sources at the earth's surface than is the observed one.

One may think of different ways of determining the parameters τ_0 and k. It would seem simplest to determine these from the observed temperatures at different heights, but temperature normals for several levels, having a hemispheric coverage, are not available. On the other hand, the tabulated height normals recently published by the U.S. Weather Bureau can readily be used to determine these parameters. The normals quoted contain 700 and 500 mb heights together with the thickness 1,000—700 mb. We now require that the temperature curve represented by equ. (2.1) gives the same thicknesses of the layers 1,000—700 mb and 700—500 mb as those observed. Moreover, we will for practical purposes have τ_0 refer to the 1,000 mb height rather than sea level; this implies that we from now on refer to the 1,000 mb surface as a zero level. As will now be seen, this requirement makes it possible to determine τ_0 and k uniquely from the thickness normals 1,000—700 mb and 700—500 mb. The general analytical expression for the thickness $z_2—z_1$ of a layer bounded by the two pressure surfaces p_1 and p_2, is

$$z_2 - z_1 = \frac{R}{g}\int_{p_2}^{p_1} T\frac{dp}{p} \qquad (2.2)$$

where T is the observed temperature, R the gas constant for atmospheric air, and g the acceleration of gravity. In order to carry out the integration of (2.2), one needs to know T as a function of p. This is obtained in the model by substituting $-\frac{dT}{k}$ (in virtue of (2.1)) for dz into the hydrostatic equation in the form

$$\frac{dp}{p} = -\frac{g}{R}\frac{dz}{T} \qquad (2.3)$$

and integrating, whereupon the following expression is arrived at:

$$\tau = \tau_1\left(\frac{p}{p_1}\right)^{\frac{Rk}{g}} \qquad (2.4)$$

Carrying out the integration of (2.2) by substituting the expression (2.4) for the temperature we obtain the following expression for $z_2—z_1$

$$z_2 - z_1 = \frac{\tau_1}{k}\left[1 - \left(\frac{p_2}{p_1}\right)^{\frac{Rk}{g}}\right] \qquad (2.5)$$

We now apply (2.5) to the two layers, first of which is bounded by p_0 (= 1000 mb) and p_1 (= 700 mb), and the second one by p_0 and p_2 (= 500 mb). The factor $\frac{\tau_1}{k}$ in (2.5) is then in both cases to be replaced by $\frac{\tau_0}{k}$ and can be eliminated by division. By doing so, we arrive at the following equation for the determination of k:

$$z_1 - z_0 = \frac{1 - \left(\frac{p_1}{p_0}\right)^{\frac{Rk}{g}}}{1 - \left(\frac{p_2}{p_0}\right)^{\frac{Rk}{g}}}(z_2 - z_0) \qquad (2.6)$$

which is easily solved graphically. (The k-lines are straight lines in a $(z_1—z_0)$, $(z_2—z_0)$ diagram). Having determined k, we then find the other parameter τ_0 (= temperature at 1,000 mb) from (2.5), which can either be applied to the layer 1,000—700 mb or the layer 1,000—500 mb. This equation has also a simple graphical solution. The graphs for the

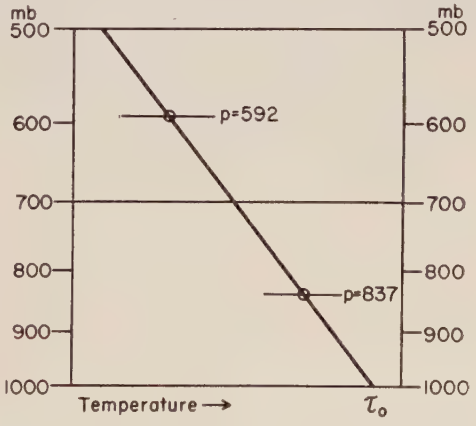


Fig. 1. The temperature-height curve of the model. For explanation see the text.

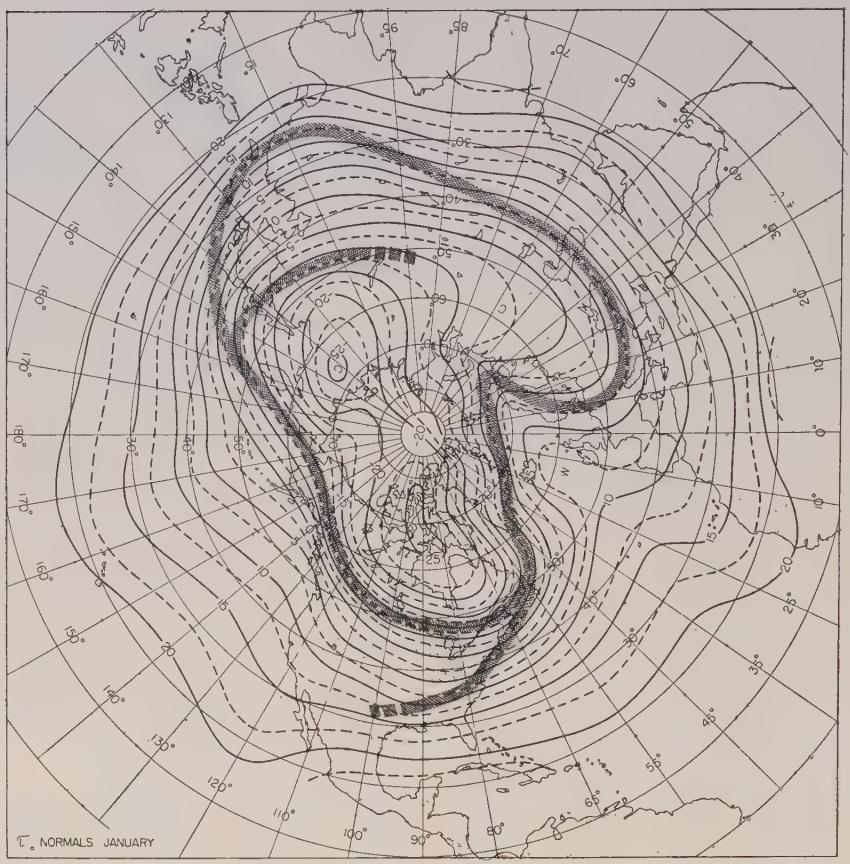


Fig. 2. The distribution of τ_0 for January.

determination of k and τ_0 have not been reproduced here, since their construction is straightforward.

The principle of the method is most easily illustrated if one refers to a p-system (p or a function of p independent variable instead of z). Fig. 1 shows a log p—T diagram where the points corresponding to ½ (log 1,000 + log 700), ½ (log 700 + log 500) and the mean temperatures for the layers 1,000—700 mb and 700—500 mb respectively have been plotted. A straight line through both of these points represents a sounding curve which gives the same thickness for both layers as the observed one. The 1,000 mb temperature, τ_0, is immediately obtained by intersection with the 1,000 mb line and the lapse-rate is determined by the slope. Since in the model we are dealing with z as the independent variable rather than log p, the straight line in the figure is not identical with the height-temperature curve defined above. For the purpose of illustrating the method, however, this difference is unimportant.

3. The τ_0-distribution

Figs. 2 and 3 show the distribution of τ_0 for January and July respectively. It follows from the definition of τ_0 (see fig. 1) that the general pattern of the τ_0-charts can not deviate too much from the pattern of the mean 1,000—700 mb thickness isopleths. Were it not for the non-uniformity in the temperature lapse-rate, the two fields would be identical. Comparison shows that by and large the features are much the same, but it shows also, that in certain regions there are significant differences. Such regions are primarily found along the border lines between the continents and the

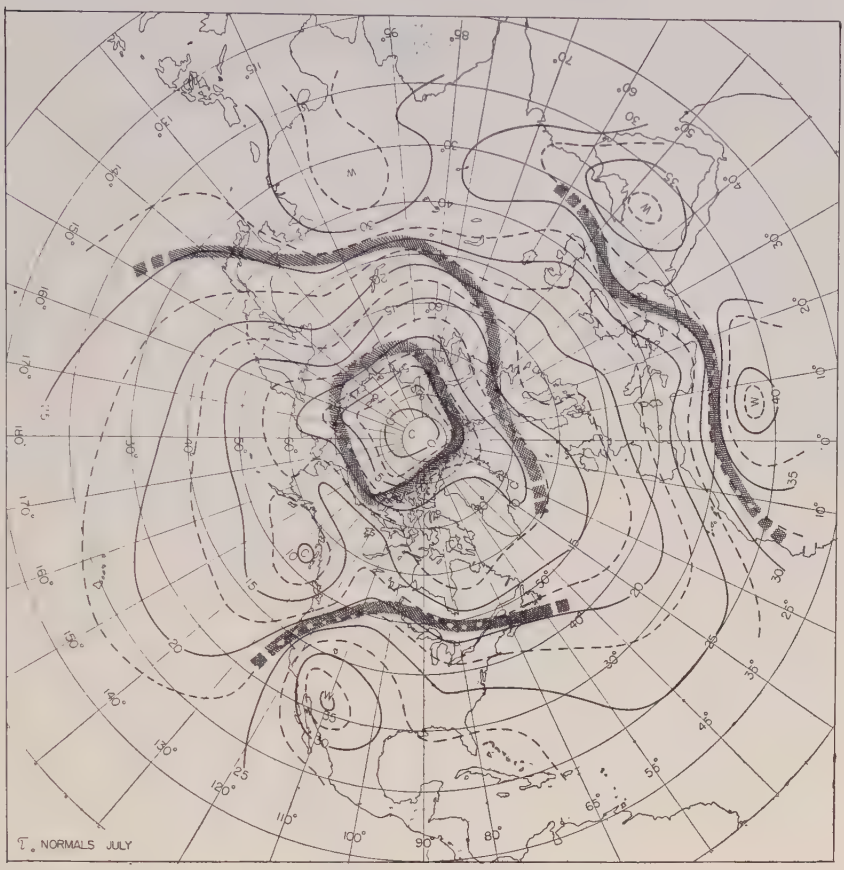


Fig. 3. The distribution of τ_0 for July.

oceans, where particularly strong lapse-rate gradients occur. As appears from the k-charts (see figs. 6 and 7) the cold air is in general more stable than the warm air, and this is particularly the case where the lowest and highest temperatures occur. This is very typical over Siberia and Canada in winter, and over the subtropical parts of the continents in summer. As a result of this, the regions of strongest 1,000—700 mb thickness contrasts will show still more pronounced contrasts on the τ_0-charts, and the concept of a front at sea level is therefore more readily associated with the latter representation of the low-level temperature distribution. In this respect the τ_0-distribution resembles that of the observed sea level and surface temperatures (Handbook of Meteorology, 1945, page 949 and 952), but will on the other hand not show the irregularities of the latter. For this reason we may refer to τ_0 as the "representative" 1,000 mb temperature. It can be shown that an ideal frontal surface would in the τ_0-representation be depicted as a frontal zone just north of the surface front; the width of this zone will depend on the slope of the frontal surface. In the following description of the zones of maximum τ_0-gradients and their relation to the mean positions of the major fronts, this should be born in mind.

The double-hatched zones in figs. 2 and 3 indicate relative maxima in the τ_0-gradients, i.e. zones of maximum low-level baroclinity. One may expect that the positions of these zones are closely related to the positions of the main surface fronts. In his book, "Weather Analysis and Forecasting", figs. 124 and 126, PETTERSSEN presents two hemispheric charts showing the mean position of the principal frontal zones at sea level in winter and

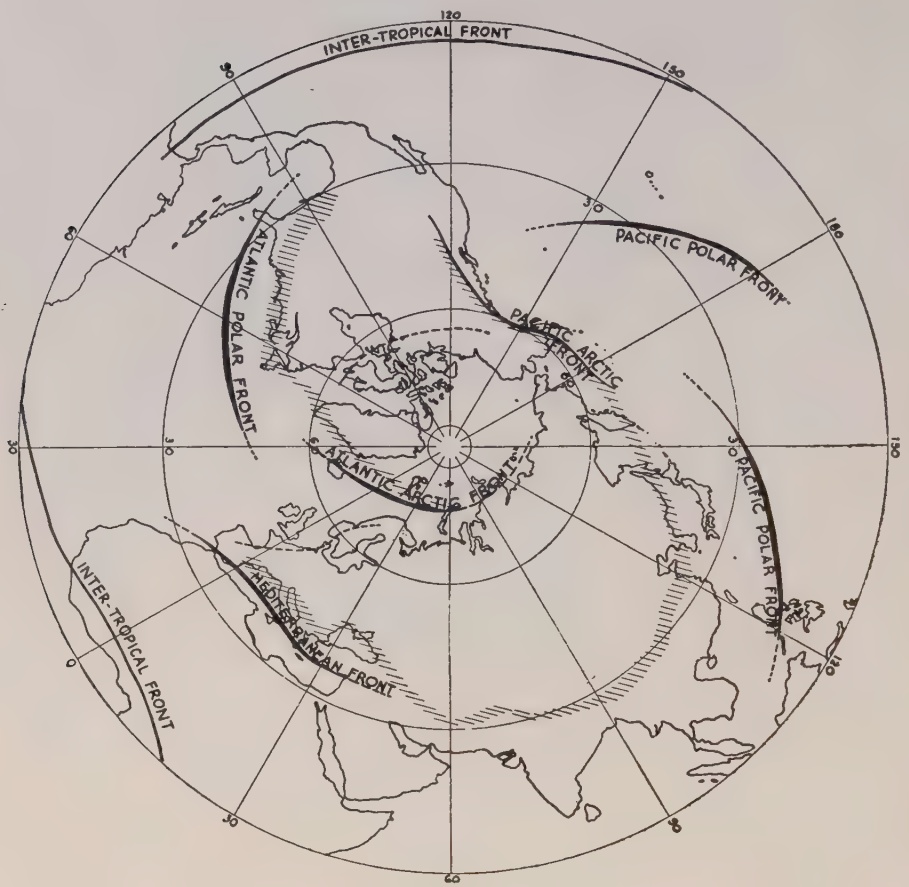


Fig. 4. The principal frontal zones in winter. Hatched zones indicate zones of maximum temperature gradient. (By permission from "Weather Analysis and Forecasting" by S. Petterssen. Copyright May 1954, McGraw-Hill Book Comp., Inc., New York.)

summer, as well as the zones of maximum temperature gradients. These maps have by due permission been reproduced as figs. 4 and 5 in this paper. It is assumed here that the zone of maximum temperature gradients shown in Petterssen's charts refer to low levels, although not specifically stated in his book. Hence, a comparison will be made between the τ_0-distribution for January and July and Petterssen's winter and summer charts respectively, beginning with the winter.

As a whole there is a good agreement between the zones of maximum baroclinity, in particular those connected with the polar and arctic fronts of the Atlantic and Pacific Oceans. The east-west branch of the maximum τ_0-gradient extending from the main zone of the North Pacific and running into Siberia has no corresponding feature in Petterssen's winter chart and is obviously not directly related to any of the principal frontal zones. On the other hand, it is quite clear from the mean sea level pressure distribution for January that the low-level advection favors a concentration of isotherms in this region. As regards the position of the zone of maximum baroclinity over Southern Europe and Asia between the longitudes 15°—80° E, the maximum in the τ_0-gradient is so flat that the hatched zone could have been placed a few degrees of latitude further to the south without violating the data, in which case there would have been a closer agreement between the two charts. The relatively weak maximum in the τ_0-gradient over Western Europe connecting the much stronger maxima

Fig. 5. The principal frontal zones in summer. (By permission from "Weather Analysis and Forecasting" by S. Petterssen. Copyright May 1954, McGraw-Hill Book Comp., Inc., New York.)

over the Arctic Ocean and Southern Europe, is not shown directly by Petterssen, but corresponds to a mean position of a front of secondary importance shown on his map. PALMÉN (1951) has pointed out that over Europe in winter there is often a tendency toward the formation of a surface front between the maritime polar air masses over the western part and the continental polar air over the eastern part of the continent. A glance at the mean sea level pressure chart for January shows that the hatched zone over Central and Southern Europe coincides with the southwestern edge of the continental Eurasian anticyclone.

The particularly pronounced maximum in the τ_0-gradient across the entire North-American continent joining the North-Pacific zone is not shown by Petterssen except for a weak front of secondary importance. The mean sea level distribution over North America shows that the zone of maximum temperature contrasts coincides with the pressure trough between the arctic high and the sub-polar continental one. Palmén has pointed out that the splitting of the continental anticyclone into two parts separated by a secondary front (between the arctic or very cold polar air on one side and the considerably warmer polar air on the other) is a common phenomenon over the United States.

Although we are primarily dealing with low-level phenomena, the τ_0-distribution will in virtue of the thermal wind relationship reflect to a certain extent the wind conditions at high levels. Thus, the two zones of maximum τ_0-gradients over Asia and North America would seem to correspond roughly

to the two main jets found on mean meridional cross sections for the winter months. Reference is made to WILLETT's (1944) well known cross section Swan Island—Bismarck—Barrow during winter showing a subtropical jet between Miami and Nashville (around 28° lat.) and a polar jet over Alaska, and also to MOHRI's (1953) winter cross section along the 140° E meridian which besides the subtropical jet (main jet) also indicates a weak polar jet at latitude 41°.

The τ_0-distribution in July is shown in fig. 3. As one may expect the gradients are much weaker than in winter and their maxima less pronounced. By and large the zones of maximum temperature contrasts coincide with the positions of the principal frontal zones shown by Petterssen (fig. 5). The extension of the Arctic Ocean band of maximum baroclinity across Siberia, China and Southern Japan may not correspond to any major front. On the other hand, such a maximum, although weak, is clearly indicated on the τ_0-distribution, and according to THOMPSON (1951) the West Pacific polar front should be found along the Tibetan Plateau, whereas Petterssen on his summer chart shows no polar front at all over the middle and southern parts of the Asiatic continent.

A feature of particular interest and not shown so clearly in any earlier presentation of the low-level temperature field is the closed excentric zone of maximum baroclinity around the polar cap. In this the northernmost part of the hemisphere we find the largest low-level temperature contrasts. In this connection we refer to SVERDRUP's (1953) experience about the atmospheric disturbances over northeastern Siberia and the Polar Sea during the "Maud"-expedition 1918—1925. According to him, the anticyclonic type of weather predominated in this area during October—April. In May a transition to summerlike conditions appeared as the weather became of more cyclonic character. "In summer several disturbances apparently were formed along a quasistationary front, which was running approximately east-west and was colder on the northern side." In September a transition to winter conditions appeared. It is likely that the circumpolar zone of maximum τ_0-gradient on the July map corresponds to the frontal zone mentioned by Sverdrup.

In much lower latitudes there is another well marked zone of maximum temperature gradient extending from the coast of West Africa to the interior of Asia north of the Gulf of Persia. Its position is just north of the main desert regions and the crowding of the τ_0-lines simply reflects the enormous heating of the Sahara desert as well as the deserts of western Asia. The westernmost part of this zone corresponds roughly to the boundary between the northeast trades and the West-african monsoon and is indicated on Petterssen's summer map as a front of secondary importance. One might expect some cyclone activity along the above mentioned zone, but according to BIEL (1944) fronts and cyclones are rare in the Mediterranean during the summer. The overall lack of rain and cloudiness in this region is of course closely related to the dryness of the air masses involved.

Zonal mean values for τ_0 are shown in fig. 10 as a function of latitude.

4. The *k*-distribution

Figs. 6 and 7 show the *k*-isopleths for January and July respectively drawn at intervals of 0.5° C km^{-1}. The letters S and I indicate respectively minima and maxima in the temperature lapse-rate. To the authors' knowledge, maps showing the normal distribution of temperature lapse-rate on a hemispheric scale have not been published earlier, and it is therefore felt that a discussion of the *k*-maps in some detail may be appropriate. The main features of these maps may, however, first be summarized as follows:

1. The pattern is more broken up than on the corresponding maps for τ_0.
2. In January the lapse-rate is in general low over the continents and high over the oceans at middle and high latitudes. These conditions are reversed in July.
3. Regions of lowest lapse-rates are northern Canada and northeastern Siberia in January and lat. 80°—90° in July. The highest values are found just south of Iceland and over the northernmost part of the Pacific in January and over the desert areas of the subtropical continents in July.
4. The strongest gradients occur along the coasts of the continents north of latitude 50° in January and north of latitude 70° in July.

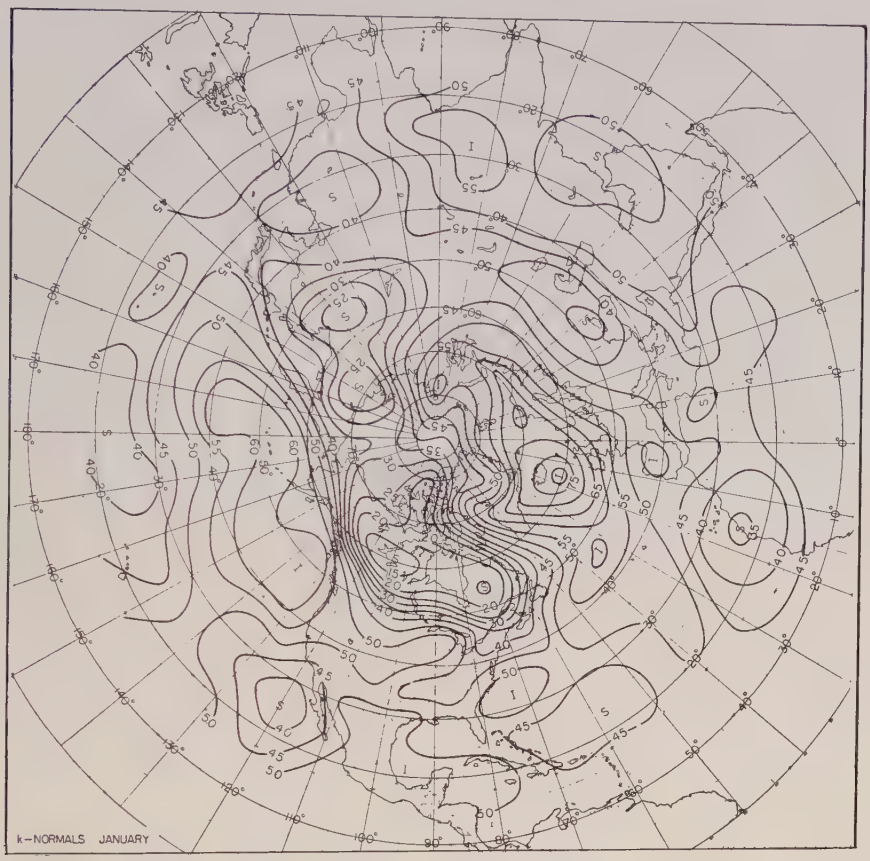


Fig. 6. The distribution of k for January. An S indicates a relative minimum and an I a relative maximum in the temperature lapse-rate.

5. There is a belt of low temperature lapse-rate around 5° of latitude south of the subtropical high pressure girdle at sea level.

The facts listed above are qualitatively well known from physical considerations and synoptic experience, although not demonstrated quantitatively to the same extent earlier. The main features of the geographical-seasonal distribution of temperature lapse-rate is, of course, essentially explained by the geographical location of heat and cold sources and their seasonal shifting. A part of the explanation is also the overall subsidence in the cold continental air and the compensating convergence over the oceans in winter, and the convergence over the continents and the divergence over the oceans in summer.

Although recognized qualitatively, the quantitative measure of some of the main features of the k-distribution may seem surprising. This is to some extent the case of the extremely strong gradients in winter between Iceland and Canada as well as the zone of strong gradients running across the North American continent through Alaska and into the North Pacific. As shown in the last section, these strong k-gradients cause substantial change with height in the thermal wind. More remarkable is perhaps the very strong k-field over the Arctic in July. The magnitude of the k-gradients in this region is comparable with the strongest gradients occurring during January. The occurrence of considerable contrasts in temperature lapse-rate between the ice-covered polar cap and the extensively heated continental surroundings in summer is a priori plausible, since the temperature of the ice can not exceed 0° C, but the strength of this field is in general not realized. It must,

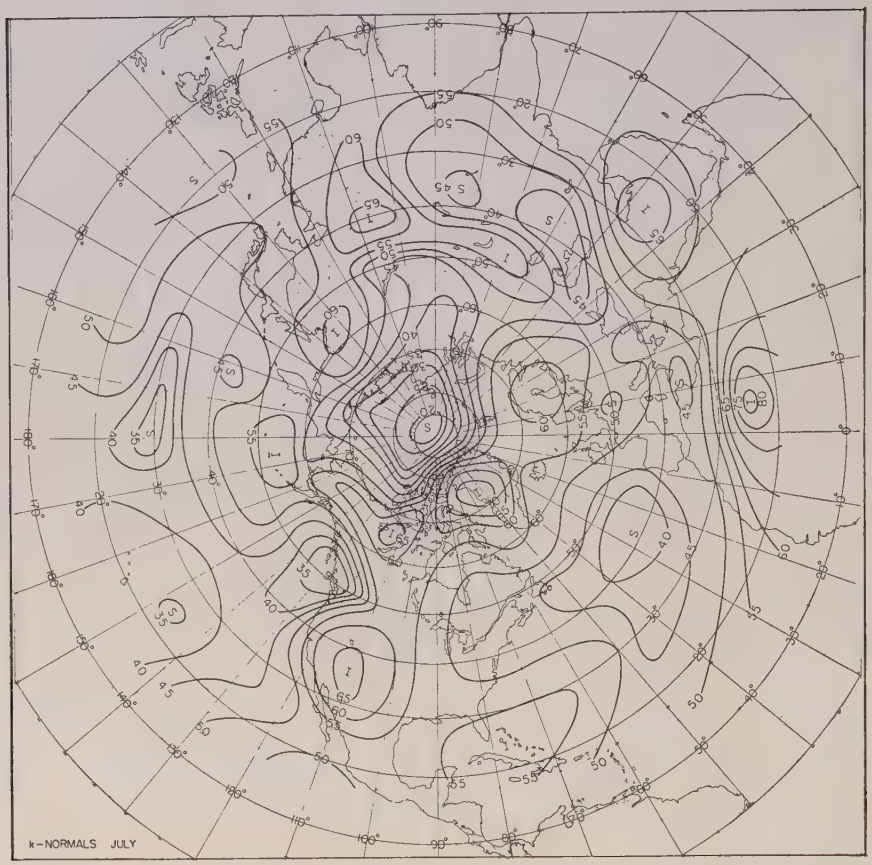


Fig. 7. The distribution of k for July. An S indicates a relative minimum and an I a relative maximum in the temperature lapse-rate.

of course, be born in mind that the analysis over the high Arctic is not as reliable as further south, and for this reason too much emphasis cannot be put upon the details. On the other hand, it is known that the height and thickness normals, upon which the k-charts are based, were carefully tested with regard to consistency of the temperature field; it therefore seems reasonable to have considerable faith in the main features of the k-distribution even at these high latitudes. A few spot checks have been possibly by comparing mean winter and summer soundings compiled by FLOHN (1950) for the arctic stations Tromsö (69°5 N, 19° E), Spitzbergen (78° N, 13°5 E, summer only), Point Barrow (71° N, 156° W), Franz Josef's Land (80° N, 53° E, winter only) and Jakutsk (62° N, 129° E, winter only). These soundings were in good agreement with the k-charts, as far as could be judged by comparing lapse-rate values for two extreme months with seasonal averages. As to the details in January map, the following is pointed out:

a. The trough of low lapse-rate extending from the interior of Canada to the southern part of the United States may correspond to the southern edge of the sinking polar air, whereas the relatively high lapse-rate between Bermuda and the coast of U.S. may reflect the heating of the cold continental air when flowing out over the warm water of the Gulf Stream. The synoptic conditions are similar over the easternmost part and the coastal water of the Asiatic continent, and the details of the k-charts are much alike in these two regions.

b. The area of large lapse-rates over the northern part of the Atlantic with Iceland as a center, is essentially explained by the frequent outbreaks of cold continental air from Canada

and the Greenland area. These outbreaks have long overwater trajectories before reaching the middle and eastern part of the Atlantic and may therefore be expected to have become fairly unstable. More direct outbreaks between Norway and Greenland are not uncommon and in spite of the shorter trajectory the relatively warm water over which the air flows will cause a considerable increase in the lapse-rate. The relation between decrease in stability and the distance travelled by cold air from the North American continent has been investigated by BERSON and PETTERSSEN (1943). They found that atmospherics might occur in the air mass at a distance exceeding 1,000 st. miles from the coast. They also showed a rapid decrease in the frequency of atmospherics south of about 45° N.

As to the details of the July map the following is mentioned:

a. There is a pronounced maximum in the lapse-rate around 65° N, i.e. approximately 5° of latitude north of the low-pressure belt (see fig. 10b) at sea level. The high lapse-rate is apparently closely connected with the frequent cyclonic activity in these latitudes during summer.

b. The western parts of the oceans show larger lapse-rates than the eastern. This agrees well with the synoptic fact that the tropical maritime air is less stable in the western halves of the anticyclonic cells than in the eastern halves.

So far it has not been mentioned that the Normal Weather Charts (1952) give observed 700 mb temperatures tabulated in the same way as the height and thickness values utilized here. In reference to fig. 1, this implies that one additional point of the temperature height curve is known besides those marked in the figure. Obviously this extra piece of information provides an additional knowledge of the stability conditions of the 1,000—500 mb layer and makes it to a certain extent possible to distinguish between the two sub-layers 1,000—700 mb, and 700—500 mb. In order to do so, the observed 700 mb temperatures have been subtracted from the corresponding 700 mb temperatures computed by means of equ. (2.1), using the appropriate τ_0- and k-values. In general the differences will not be zero, and they are obviously a rough measure of the deviation in the lapse-rates of the two sub-layers from the k-values. A negative value of a difference means in general that the lapse-rate of the layer 1,000—700 mb is less than the lapse-rate given by the k-value, whereas the layer 700—500 mb has a lapse-rate exceeding the k-value, i.e. the lowest layer is the more stable one. A positive value on the other hand indicates that the uppermost layer is the more stable one. In converting the differences of the two sets of 700 mb temperatures into differences in lapse-rate, we may for practical purposes assume the thicknesses of both layers to be the same and equal to 2.5 km. This means that a difference of 1° C corresponds to a deviation of 0.4° C km^{-1} in the lapse-rate of each layer from the appropriate k-value, i.e. the lapse-rates of the 1,000—700 mb and 700—500 mb layers differ 0.8° C km^{-1}.

A study of these 700 mb temperature differences gave the following results:

1. On the average the lapse-rate of the layer 1,000—700 mb is slightly less than the lapse-rate of the layer 700—500 mb. This difference is very small in July, whereas it in January is well marked in the latitudes 40°—70° having a maximum of 1.6° C km^{-1} at 60° (each of the layers deviates 0.8° C km^{-1} from the k-value). South of latitude 30° and north of latitude 80° the conditions seem to be reversed both in July and January, i.e. the uppermost layer is the more stable one. The difference is less than 0.4 in the low latitudes but varies from 0.4—1.0 in the high latitude belt. Here, of course, one must bear in mind the uncertainty of data near the north pole as well as in the low latitudes.

2. There is a pronounced seasonal variation over the continents at middle and high latitudes. In January the layer 1,000—700 mb is by far the more stable one, this is particularly the case between the cores of the Canadian high and the Icelandic low and also over northeastern Siberia between the continental high and the Aleutian low. The maximum lapse-rate difference between the two layers is 3.2° C km^{-1}, but does in most places not exceed 1.6° C km^{-1} in the regions mentioned. In July the conditions tend to be the opposite over the continent, but in general the difference in lapse-rate between the two layers is considerably less than in January. Maximum values of 2.4° C km^{-1} occur over the interior

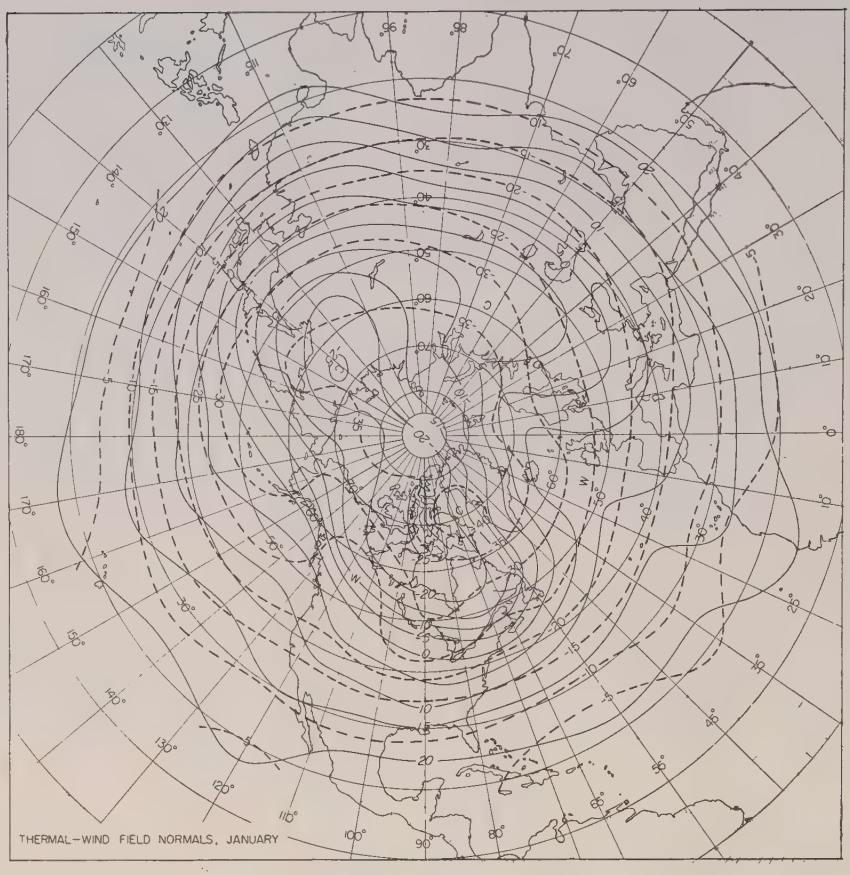


Fig. 8. The thermal wind distribution at 1,000 mb (solid lines) and 500 mb (broken lines) for January. For further explanation see the text.

of Asia north of lat. 45° and over the northwesternmost part of the U.S.A.

3. In addition to the overall picture given above, there are some details deserving of mention. Such details are the two comparatively small areas of positive differences in the 700 mb temperatures (1,000—700 mb layer the less stable one) in January off the east coast of Japan and over the Atlantic south of Greenland and east of Newfoundland. The first of these two is the more pronounced. One is inclined to explain the existence of these two regions as a result of the rapid heating of the cold continental air when it travels over the much warmer ocean as discussed earlier. Another feature of interest in January is an oblong area at around lat. 75°N running from the east coast of Greenland to longitude 75°E where the layer 1,000—700 mb is considerably less stable than the layer 700—500 mb. Apparently these large low-level lapse-rates in this region are to some extent a result of the considerable cyclone activity there, but may also reflect a certain lack of thermal consistency in the analysis of the normal height maps. Noteworthy features in July are that whereas the Azores high hardly shows any preference for higher stability in the lower layer, the Pacific high is very much more stable below 700 mb than above. It is finally mentioned that north of lat. 80° there are indications that the layer 700—500 mb is the more stable one during July.

Zonal mean values for k are shown in fig. 10 as a function of latitude.

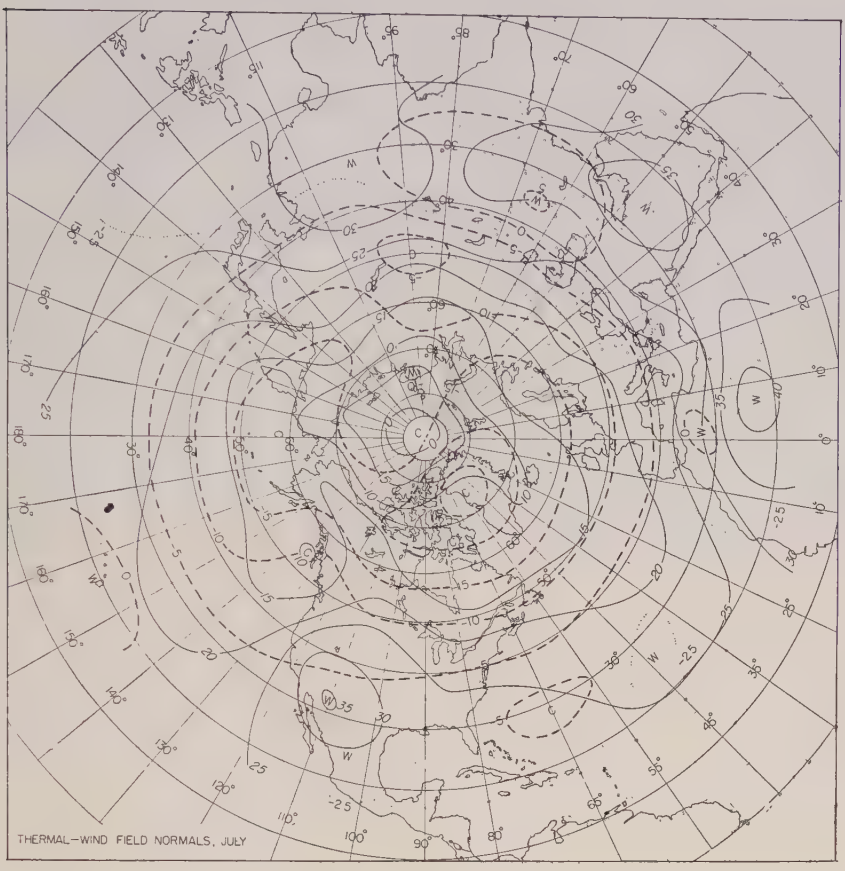


Fig. 9. The thermal wind distribution at 1,000 mb (solid lines) and 500 mb (broken lines) for July. For further explanation see the text.

5. The thermal-wind field

If $\mathbf{v}$ denotes the geostrophic wind, $\frac{\partial \mathbf{v}}{\partial z}$ is the thermal wind, commonly expressed in terms of the horizontal temperature ascendent as follows:

$$\frac{\partial \mathbf{v}}{\partial z} = \frac{g}{fT} \mathbf{k} \times \nabla T \qquad (5.1)$$

where $\mathbf{k}$ is a unit vector directed vertically upward and f is the Coriolis parameter. We now substitute in equ. (5.1) the temperature of the model, as given by equ. (2.1), and obtain the following expression for the thermal wind

$$\frac{\partial \mathbf{v}}{\partial z} = \frac{g}{f\tau} \mathbf{k} \times \nabla \tau = \mathbf{v}' - z\mathbf{v}'' \qquad (5.2)$$

where

$$\mathbf{v}' = \frac{g}{f\tau} \mathbf{k} \times \nabla \tau_0; \quad \mathbf{v}'' = \frac{g}{f\tau} \mathbf{k} \times \nabla k \qquad (5.3)$$

The variation with height in $\frac{1}{\tau}$ is unimportant, and we may therefore for practical purposes consider the vectors $\mathbf{v}'$ and $\mathbf{v}''$ as independent of z. Consequently, the thermal wind varies linearly with height. Since there is strong empirical evidence for the linearity with height in the temperature, in particular when averaged in time, we believe that equ. (5.2) gives an essentially correct picture of the normal distribution of the thermal wind. It is pointed out that in spite of the simple thermal distribution, the model allows not only for a

change in magnitude, but also for a change in direction of the thermal wind with height.

It appears from equ. (5.3) that the vectors $\mathbf{v}'$ and $\mathbf{v}''$ are related to the parameters τ_0 and k in a similar way as the geostrophic wind is related to the contour lines; obviously $\mathbf{v}'$ is the thermal wind at $z=0$ ($=1{,}000$ mb), whereas $\mathbf{v}''$ expresses the constant rate of change with height. It now follows that the τ_0-and k-maps together depict the field of thermal wind. Since the isotherms are drawn for every fifth degree centigrade, and the lapse-rate isopleths for an interval of $5 \cdot 10^{-4}$ °C m^{-1}, the two right hand terms in equ. (5.2) are equal in magnitude at $z = 5$ km wherever the k-lines are twice as dense as the τ_0-lines. Having this in mind, an inspection of the two maps immediately reveals that in certain regions there is a very considerable variation in the thermal wind with height, particularly in winter.

In order to show this variation more clearly, the temperature τ_5 at the 500 mb level has been computed by means of equ. (2.1) and its isopleths at 5°C interval superimposed upon the τ_0-maps; those are shown in figs. 8 and 9. The solid lines are the τ_0-isopleths and the dashed lines the isotherms of the 500 mb surface. An equal spacing of the two sets of curves corresponds to the same magnitude of the thermal wind, and its direction is along the respective isotherms with the lowest temperature to the left. Although these two maps primarily show the isobaric distribution of temperature, it is convenient to refer to them as thermal wind maps in virtue of the relationship expressed by the equations (5.2) and (5.3). By the aid of the thermal wind charts, one can easily study the change with height of the thermal wind.

The main features of the thermal wind distribution with increasing height may be summarized as follows:

1. Pronounced change both in magnitude and direction at middle and high latitudes.
2. General tendency toward more regular pattern and toward predominance of zonal direction.
3. Strong variation in regions of pronounced low-level temperature advection. Typical regions of this kind in winter are the Newfoundland—Greenland—Iceland area and Western Europe.

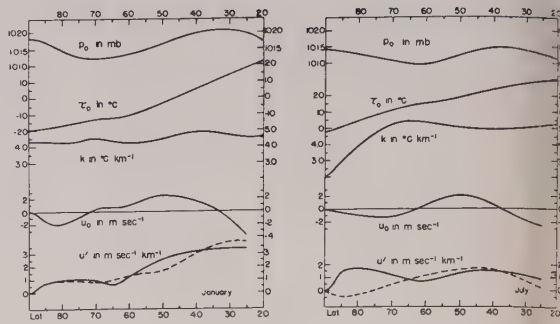


Fig. 10. Zonal means for the sea level pressure p_0 and the corresponding geostrophic wind u_0, the parameters τ_0 and k as well as the thermal wind u' at 1,000 mb (solid lines) and 500 mb (broken lines), a) for January, b) for July.

4. Extremely great decrease in magnitude in January over the northeasternmost part of Siberia, Alaska and western and southern Canada.
5. Marked decrease over the Arctic in July.

The zonal mean of the thermal wind, denoted by u' is shown in fig. 10 as a function of latitude. The solid curves refer to the 1,000 mb and the dashed ones to the 500 mb level. For the computation of u', central differences in the zonal means of the respective temperatures comprising ten degrees of latitude were used. One notices the minima in the thermal wind at 1,000 mb at lat. 65° and 60° in January and July respectively, which reflect the overall weak low-level temperature gradient in these latitudes. Noteworthy is the strong thermal wind at 1,000 mb in high latitudes in July and its rapid decrease with height, resulting in the reverse direction at 500 mb north of latitude 75°. There are presumably some doubts as to the reality of this reversal. In the first place, the upper air analysis north of latitude 60° but south of 75° is based upon a sparse network of stations and in the second, the analysis north of latitude 75° is practically a mere extrapolation of the contour distribution to the south.

Acknowledgement

The authors wish to express their thanks to the technical staff of the Institute of Meteorology, University of Stockholm; in particular to Miss B. WICKBERG who drafted the figures and Mrs. U. TABATABAI who carried out most of the computations for the preparation of the charts.

REFERENCES

Árnason, G., 1952: A baroclinic model of the atmosphere applicable to the problem of numerical forecasting in three dimensions. I. *Tellus*, 4, 356—373.

—, 1953: A baroclinic etc. II. *Ibid.*, 5, 386—402.

Berry, F. A., et al., 1945: *Handbook of Meteorology*, New York, McGraw-Hill Comp., pp. 949 and 952.

Berson, F. A., and Petterssen, S., 1943: Atmospherics in relation to fronts and air masses. A report of *Meteorological Research Committee, Air Ministry*, London, 7 pp.

Biel, E. R., 1944: Climatology of the Mediterranean area. Univ. of Chicago, Dep. of Meteorology, *Miscellaneous Reports* No. 13, pp. 12 and 135.

Flohn, H., 1950: Grundzüge der allgemeinen atmosphärischen Zirkulation auf der Südhalbkugel. *Arch. Met. Geophys. und Bioklimat., A*, 2, 17—64.

Mohri, K., 1953: On the fields of wind and temperature over Japan and adjacent waters during winter of 1950—1951. *Tellus*, 5, 340—358.

Palmén, E., 1951: The aerology of extratropical disturbances. *Compendium of Meteorology*, Amer. Met. Soc., Boston, Mass., pp. 599—620.

Petterssen, S., 1940: *Weather Analysis and Forecasting.* New York, McGraw-Hill Comp., pp. 269 and 271.

Sverdrup, H. U., 1953: *North Polar Expedition with the "Maud", 1918—1925.* Vol. 2, Part 1. Bergen, A.S.J. Griegs Boktrykkeri, pp. 324—325 and 328.

Thompson, B. W., 1951: An essay on the general circulation of the atmosphere over South-East Asia and the West Pacific. *Quart. J. R. Met. Soc.*, 77, 569—597.

U.S. Dep. of Commerce, Weather Bureau, 1952: *Normal Weather Charts for the Northern Hemisphere.* Technical Paper No. 21. Washington, D. C., Government Printing Office, 74 pp.

Willett, H. C., 1944: *Descriptive Meteorology.* New York, Academic Press, Inc., p. 132.

On the Correlation between Geomagnetic Activity and the Diurnal Variation of Cosmic Rays

By ARNE ELD SANDSTRÖM, Fysiska Institutionen, Uppsala

(Manuscript received December 9, 1954)

Abstract

The cosmic ray measurements of K.-G. Malmfors (1949) Oct. 1947—May 1950 have been studied with regard to the effect of geomagnetic activity on the diurnal variation. By means of the same data a study was also made of the secular phase shift first noticed by Elliot and Thambyaphillai (1953). The time of maximum has been found to decrease with increasing values of the international K_p-index, while the amplitude itself increases. These effects were studied for the registration period as a whole as well as for short intervals. The statistical fluctuations are too big for the study of intervals shorter than half a year. The secular phase shift is the same regardless of direction. Also the phase shift accompanying geomagnetic activity is independent of the direction of measurement. This is a good proof that both shifts are due to changes in the original direction of the primaries. It is also shown, that the secular phase shift is independent of the daily geomagnetic disturbances. Apparently two effects exist, both resulting in a change of the direction of the primary cosmic ray particles.

1. Introduction

In Oct. 1947 K.-G. Malmfors started measuring the cosmic ray intensity in the Z, N, and S directions in Stockholm (magn. lat. 58°). Some of these measurements (Oct. '47—Jan. '49) have been published as a study of the daily variation (Malmfors 1949). The counter telescope arrangement is described in the same paper. As regards the present study, the following features are of interest. The N and S directions made angles of 27° with the Z direction. The maximum angular aperture in the NS plane was roughly 50° and in the EW direction 105°. The telescopes consisted of four channels in each direction, the normal counting rate for each N and S channel being 8,000 and for each Z channel 12,000 counts an hour. The counters were arranged for twofold coincidences. In the present paper Malmfors' measurements for the period Oct. '47 to May '50 are treated with regard to the influence of geomagnetic activity. Sekido and Kodama (1952) reported the existence of such an effect.

The counting rate cited above is too low for a detailed day by day study. Accordingly, we are obliged to do the sums of the bihourly values for sequences of days having approximately the same degree of geomagnetic activity. However, if a phase shift is superimposed upon the expected effect, the latter may possibly be partly obliterated. As regards the time of maximum amplitude Elliot and Thambyaphillai (1953) have shown that such a phase shift really exists.

2. The Secular Phase Shift

The diagram of Elliot and Thambyaphillai (1953) includes one point from Malmfors' measurements. As it ought to be especially

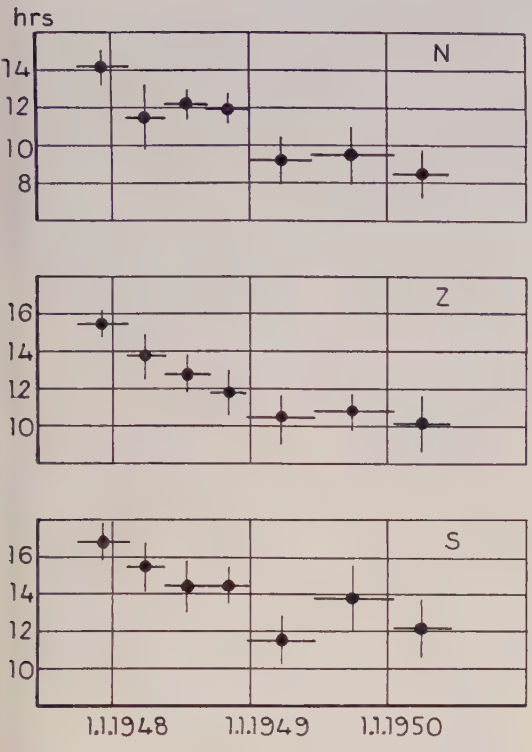


Fig. 1. The phase shift from Oct. '47 to May '50. The length of the vertical lines through the points of measurements represent the statistical fluctuations. The horisontal lines indicate the intervals within which the sums of the bihourly values were done.

interesting to follow the changes with one and the same set of telescopes, a similar diagram has been drawn using Malmfors' original "100"-day periods supplemented with three further periods (Fig. 1).

From Oct. '47 to May '50 there is a marked decrease regarding the time of maximum amplitude. Within the statistical fluctuations the shift is the same in all the three directions. From the diagrams (Fig. 1) we deduce an average continuous shift of 2 hours a year. Accordingly, if the available cosmic ray data are arranged in half year intervals, the amplitude vector will move through 1/24 of the harmonic dial during each interval. As this is a comparatively small part of a whole period, the sum of the first harmonics will certainly mirror the true features of the components. However, when the days are classified according to geomagnetic activity, days belonging to one and the same class never constitute a continuous sequence where half a year or longer intervals are concerned. Accordingly the resulting harmonic might still be influenced by the distribution of the days. Especially as regards such a long interval as Oct. '47—May '50 we recognize the possibility of a considerable influence from such sources, at least as far as classes containing a small number of days are concerned.

3. Classes of Geomagnetic Activity

The planetary index K_p was taken as a measure of the daily geomagnetic activity (Bartels 1949). This international index is tabulated for every three hours a day in tables published by the international Union of Geodesy and Geophysics (Howe and Weisman 1949, Bartels and Veldkamp 1950, 1951). The days were classified according to the highest K_p-number for any of their 3-hour periods (Table 1).

For the whole period Oct. '47—May '50 there are only 26 days in class I and 9 in class V. Accordingly the statistical fluctuations are very great, especially as regards the latter case. Some doubts may arise concerning our mode of classifying the days according to their maximum K_p-index, even if most of their 3-hour periods have low indices. It seems quite natural to use the daily sums of the

Class	K_p-index maximum value	Number of days														
		Oct. '47—March '48			Apr. '48—Sept. '48			Oct. '48—March '49			Apr. '49—May '50			Oct. '47—May '50		
		N	Z	S	N	Z	S	N	Z	S	N	Z	S	N	Z	S
I	1+	6	6	6	4	4	4	6	6	6	9	7	10	25	23	26
II	3+	71	71	71	72	71	72	50	49	50	133	78	113	326	269	306
III	5+	65	67	67	68	68	68	78	78	78	76	72	75	287	285	288
IV	7+	7	7	7	9	9	9	19	19	10	17	15	17	52	50	52
V	9+													9	9	9

K_p-indices instead. However, as yet our knowledge about the interchange between cosmic rays and geomagnetic activity is insufficient for a decision between these two types of classification. A survey also reveals that days with one or a few K_p-indices above the selected limit usually have most of their other K_p-indices just below the same limit. Accordingly, with the present rate of accuracy this question is undoubtly of slight significance. As regards all the five classes of Table 1 the average K_p-index is approximately 2/3 of the maximum value.

4. The Statistical Fluctuations

The precision of the two harmonic coefficients p and q is limited mainly by the statistical fluctuations. The corresponding influence on the value of the amplitude and the time of maximum can be calculated directly from the bihourly values. The fluctuations of the maximum amplitude are evidently limited by

$$\pm 100 \cdot \frac{|p\,\Delta p| + |q\,\Delta q|}{\sqrt{p^2+q^2}}$$

The numerical factor indicates that p and q are expressed in per cent of the average cosmic ray intensity. The statistical fluctuations affect p and q by the same amount. Let N_i be the number of impulses during each individual bihourly period and N the average. Then we have the relations

$$\Delta p = \frac{1}{6} \cdot \sqrt{\sum_{i=0}^{2\pi} \left[\Delta \frac{N_i - N}{N}\right]^2 \sin^2 \varepsilon_i}$$

$$\Delta \frac{N_i - N}{N} = 2\sqrt{N}$$

Accordingly, the value of the maximum amplitude is statistically limited by

$$\pm 100 \cdot \frac{\sqrt{6}}{3} \frac{|p| + |q|}{\sqrt{p^2+q^2}} \cdot \frac{1}{\sqrt{N}}$$

If in the harmonic dials a circle with a radius as given by this expression is drawn with the point p/q for a centre, we also get a measure as to the influence of the statistical fluctuations on the time of maximum.

The telescopes were arranged for two-fold coincidences. Accordingly, the influence of spurious counts cannot be neglected. The statistical fluctuations calculated from the total number of counts will be too low by a factor depending on the ratio of spurious counts. The increment is far from constant, as this ratio is a function of the background and therefore also of the age of the counter. However, in the present case the increase of the statistical fluctuations does not exceed 20 per cent. This estimate has not been included in the circles indicating the statistical uncertainty.

5. The Influence of Trends

The daily averages show a systematic variation with a period of one year, but the trend is not sufficiently large to affect the daily variation. The 27-day period of the daily cosmic ray intensity is more important, as are the less regular changes covering one or two days. A negative trend causes the daily maximum to appear too early, while a positive trend displaces it towards the other direction. Most important is that the amplitude is diminished by all negative trends. A large one tends to obliterate the daily variation.

However, even when the days are divided into classes according to geomagnetic activity, days with positive and negative trends will be comparatively evenly distributed. The number of days with a negative trend does not seem to be a function of the geomagnetic activity.

For a closer study the days of classes II and III during Oct. '47—March '48 were divided into two groups according to the sign of the trend. As expected, the maximum appears earlier in the day for those having a negative trend than for those having a positive trend, but the difference in time is less than the range of statistical fluctuations. Accordingly we conclude that, at least in the present case, the distribution of days between positive and negative trends is sufficiently random to eliminate the displacement of the time of maximum. As the number of days has to be sufficiently large to fulfill the condition of random distribution, the influence of the trend will be most apparent as regards days with very low or very high geomagnetic activity, i.e. classes I and V. However, in the latter case the amplitude is fairly great and the trend can therefore be correspondingly large without any serious effects.

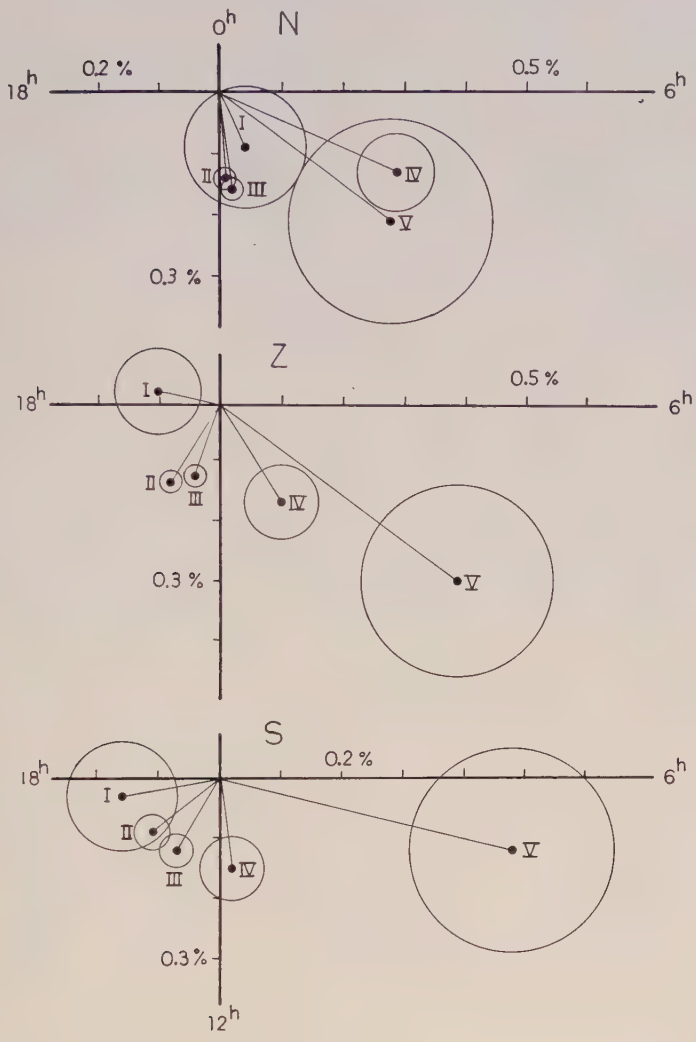


Fig. 2. Harmonic dials for the N, Z, and S directions showing the effect of geomagnetic activity. The circles indicate the magnitude of the statistical fluctuations.

6. Oct. '47 – May '50

Although the slow phase shift certainly influences the average over a long period of the time of maximum amplitude, the accompanying relative decrease of the statistical fluctuations is a sufficient reason for calculating the sums of the bihourly values for such a long interval as Oct. '47 to May '50. Dividing the days into the classes I to V (Table I) we get the results represented by the harmonic dials of Fig. 2. There cannot be any doubt whatsoever that geomagnetic activity affects the time of maximum amplitude as well as the amplitude itself. This is fully displayed by the Z and S directions. The N direction reveals to some extent different features, the effect being so small as not to surpass the limited resolution.

As regards the Z direction the maximum shift is slightly less in the present case than in that reported by SEKIDO and KODAMA (1952). These authors found a 12

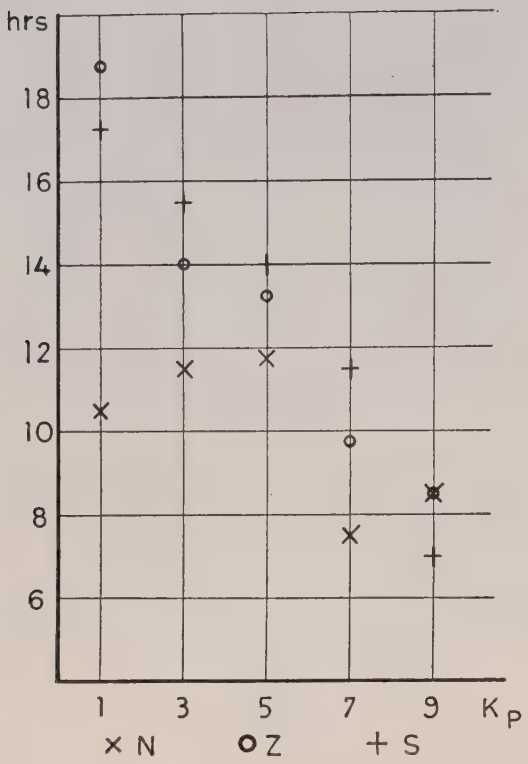


Fig. 3. The time of maximum amplitude as a function of geomagnetic activity.

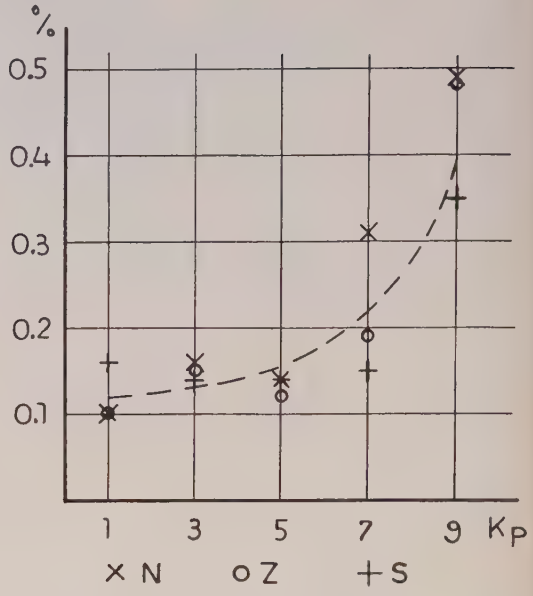


Fig. 4. The amplitude as a function of geomagnetic activity.

hours' difference where we have 10 only. With due regard for the statistical fluctuations this discrepancy is insignificant. Below we shall discuss at some length whether the phase shift is really as large as indicated even by the lower of these two values.

In Fig. 3 the time of maximum amplitude is given as a function of geomagnetic activity. The Z and S directions display the same average shift. As regards the N direction, the difference as compared to the other two directions is more apparent in this diagram than in the harmonic dials. However, we must always remember the unreliability due to the statistical fluctuations. As has been mentioned the ultimate limits of errors may be twice as great (compare Fig. 7).

Apparently the amplitude shifts in the same way regardless of direction. Fig. 4 shows that the points representing the Z, N, and S directions can be regarded as distributed along one and the same curve.

7. Short intervals

The measurements from Oct. '47 to May '50 have been divided into three six month intervals and one interval of 14 months. The latter had to be added because from April '49 until May '50 the telescope set did not work as well as before. Although a lot of valuable data were acquired, the registrations were broken off too frequently by several of the counters' becoming unreliable at one and the same time.

During each interval class V was represented by too few days to justify its being plotted in the harmonic dials. Even as regards classes I and IV the statistical fluctuations are so great as to make it impossible to resolve them from the neighbouring classes. Nevertheless we are able to decide that Fig. 5 displays the same general tendency as do the harmonic dials for Oct. '47—May '50. The small separation between classes II and III is as apparent in these diagrams as in those of Fig. 2. Considering the limits of errors, the amplitudes appear to follow the same rule as do the averages for Oct. '47 –May '50.

An attempt was also made to study the influence of geomagnetic activity during calendar months. However, from Fig. 5 it already

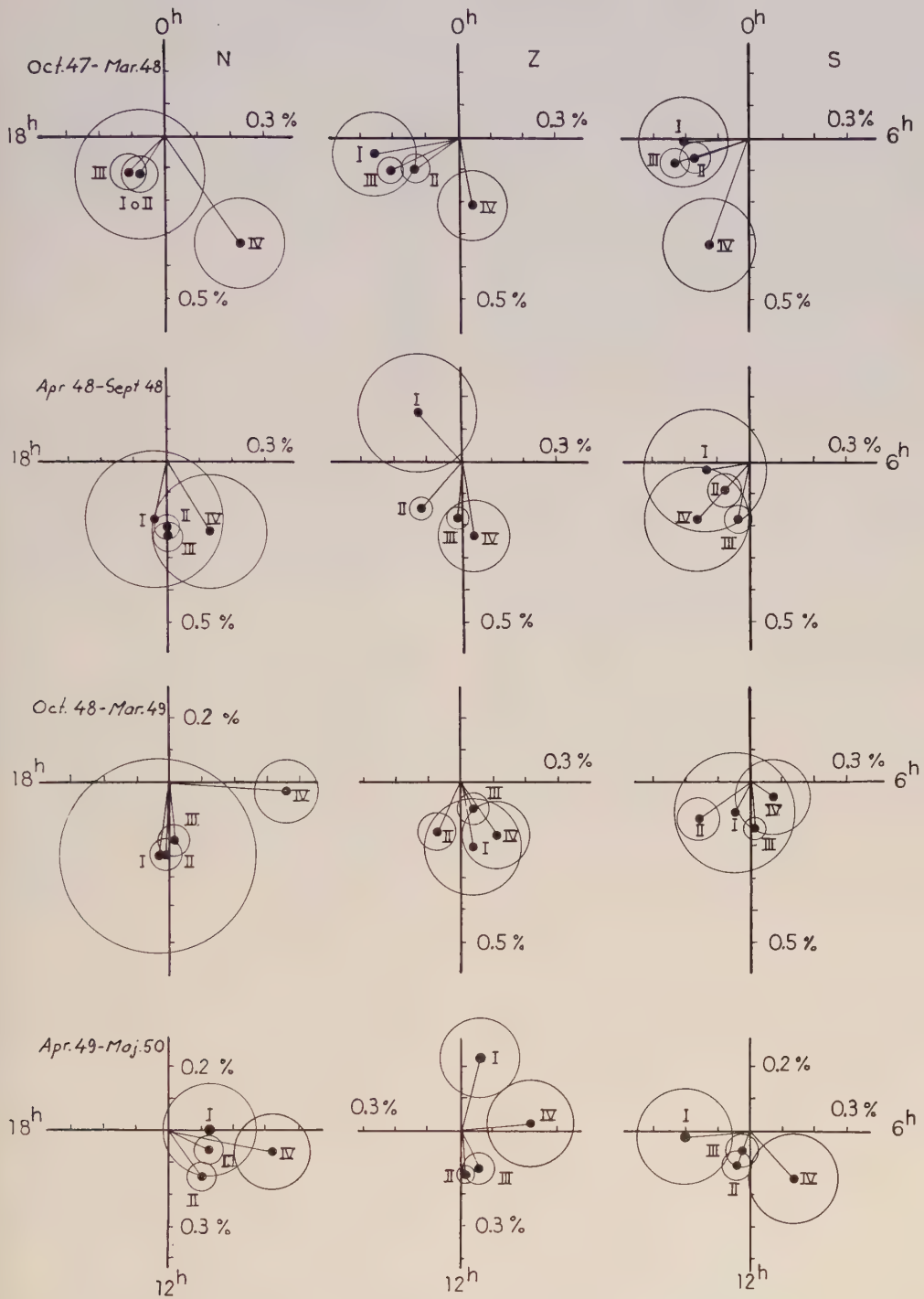


Fig. 5. Harmonic dials for the half year intervals.

appears that with a counting rate as low as in the present case half a year is probably the shortest possible interval.

According to BRUNBERG and DATTNER (1954) a geomagnetic storm probably affects the cosmic rays for a time after the end of the storm itself. As the time of maximum is a function of the geomagnetic activity, it ought to be possible to study some of the variations by doing the sums of the bihourly values for a 5—7 day interval before the period of increased geomagnetic activity, the days during the increased activity and a sequence of 5—7 day intervals during the period of decreasing activity. For this purpose we selected five periods with high geomagnetic activity preceding fairly long intervals of quiet days.

As can be gathered from the two examples in Fig. 6, the accuracy is far too low for any reliable conclusions to be drawn. In the second example the resolution is sufficiently high to reveal the usual tendency of the time of maximum to decrease with increasing geomagnetic activity. There are also some indications that the maximum appears early in the day during the first part of decreasing activity.

Mostly, however, the harmonic dials show the same features as the first example of Fig. 6, the statistical fluctuations being great enough to account for the whole shift. The counting rate ought to be so high that it becomes possible to follow the changes day by day. To be able to resolve time differences of the order of one hour, the counting rate has to be at least $5 \cdot 10^6$ counts/hour.

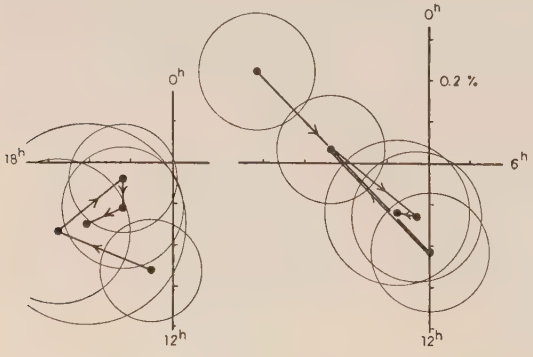


Fig. 6. Harmonic dials for the Z direction during two intervals of varying geomagnetic activity. The points represent the following intervals in order: To the left: Oct. 1—7, 10—15, 18—24, 25—30, and Oct. 31—Nov. 6, 1947. To the right: March 5—11, 12—15, 16—22, 23—29, and March 30—Apr. 5, 1948.

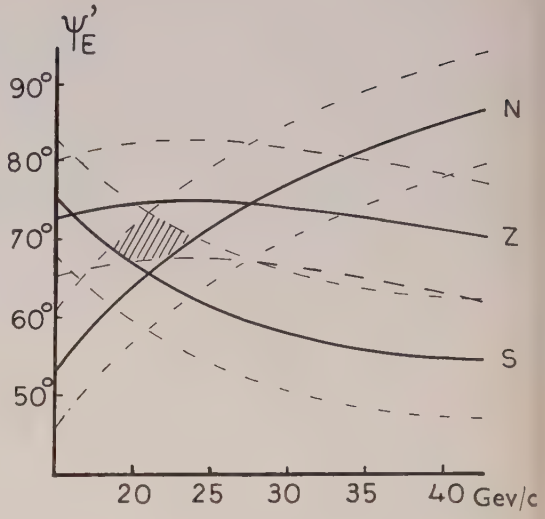


Fig. 7. The direction of anisotropy in the equatorial plane as a function of momentum according to Brunberg and Dattner (1954). The dotted lines indicate the statistical fluctuations. Ψ'_E is the sum of local time and the deflection of the particles due to the magnetic field of the earth. Geomagn. lat. 58°.

8. The Energy of the Primaries

BRUNBERG (1953) has made an experimental study of orbits of charged particles in the earth's magnetic field. As a result the true direction of the primary particles can be calculated from the directions and geometry of the counter telescopes. As a further step BRUNBERG (1953) and BRUNBERG and DATTNER (1954) calculated the east—west deflection as a function of the momentum of the primaries. The curves corresponding to the N, Z, and S directions ought to pass through one and the same point corresponding to the average energy of the primaries. Their diagram is reproduced in Fig. 7 with the addition of the statistical fluctuations. With due regard for the latter, Brunberg and Dattner found the average momentum of the primaries to be 20 GeV/c.

It is of special interest to observe how the diagram of Brunberg and Dattner is affected by a phase shift of the diurnal variation. If the shift is the same as regards all the three directions, the corresponding curves will move up or down in the diagram, the constant momentum coordinate of the intersection point indicating that the average momentum of the primaries is unaffected. Then the shift cer-

tainly originates from changes in the directions of the primaries.

On the contrary, if the shifts of the N, Z, and S directions differ, the intersection point will be displaced as regards the momentum coordinate. Accordingly there is a change of the energy distribution either as regards all the three directions or as regards one or two of them.

9. Discussion

Fig. 1 reveals the secular phase shift of the diurnal variation as being the same regardless of direction. Accordingly, from what has been said above, the energy distribution does not vary but the whole shift is due to a directional change of the primaries. This is in full agreement with the theory of ALFVÉN (1954), which tells us that the phase shift is due to the interplay between two 24 hours harmonics having a phase difference of 6 hours. The majority of cosmic ray particles is bound to the rotation of the sun. They overtake the earth in its orbit, producing a maximum at 18^h. In addition there is a direct radiation of particles from the sun. Propagated in a radial direction, these particles produce a maximum at 12^h. An early maximum indicates that the radial stream of particles dominates over that overtaking the earth in its orbit. Both streams of particles are deflected by the earth's magnetic field with a phase shift as the result. By means of the curves published by BRUNBERG (1953) the original directions of the primaries can be determined from their ultimate directions as observed at the surface of the earth. Unless there is a change of energy distribution, the shift should be the same regardless of the direction of the telescopes. As has been pointed out, this is the case as regards the measurements in Stockholm 1947—1950.

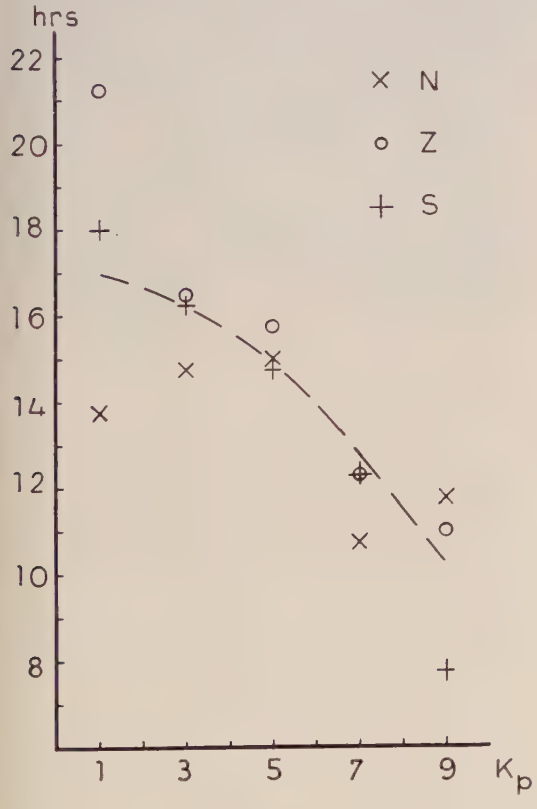


Fig. 8. The time of maximum as a function of geomagnetic activity after correcting, according to Brunberg, for the deviation of the primary particles in the magnetic field of the earth.

Turning to the geomagnetic effects our discussion meets with difficulties mostly due to the comparatively low accuracy of the telescope measurements. As already mentioned, the N direction differs in a marked manner from the other two directions (Fig. 3). If this difference were real, the N direction would display a change of energy distribution accompanying the increase of geomagnetic activity, while the other two directions would not. However, Fig. 4 indicates that inside the limits of errors the amplitude shift is the same in all the three directions. This makes it more difficult to understand why the N direction differs as regards the time of maximum. Regarding classes III to V, only, we find that the shift can be considered as the same regardless of direction.

According to BRUNBERG (1953) and BRUNBERG and DATTNER (1954) we ought to find the same time of maximum of each one of the three telescope directions, if we correct for the deviation by the magnetic field of the earth. In Fig. 8 this correction has been applied by adding $3^1/_4$ hours to the N-, $2^1/_2$ hours to the Z-, and $^3/_4$ hour to the S-direction (compare Fig. 3). In each case the points representing classes II, III, and IV fall close together. With due regard for the low accuracy, the points of classes I and V are also inside the limits of errors. The low accuracy also explains why the three directions become mixed up as

regards the time of maximum on very quiet and very disturbed days as compared to those of average geomagnetic activity.

Studying the harmonic dials in Fig. 5 we find that as regards the Z and S directions during days belonging to class I the maximum nearly coincides with that of classes II and III in five cases out of eight. As regards the Z direction we also note that one of the harmonic dials (Apr. '48—Sept. '48) disagrees thoroughly with all the others. This might cause the disagreement already remarked upon.

The general impression from the harmonic dials of Fig. 5 is that the phase shift is very small between classes I, II, and III. Then, as a function of the K_p-index, the time of maximum ought to follow the dotted line in Fig. 8.

From the discussion above the author feels justified in concluding that the observed differences between the N direction and the other two directions is due to insufficient accuracy only. Concerning classes II (maximum K_p-index 3^+), III (maximum K_p-index 5^+), IV (maximum K_p-index 7^+), and V (maximum K_p-index 9^+), the shift is undoubtedly the same regardless of the direction of the telescopes. As regards the very quiet days of class I (maximum K_p-index 1^+) there still is a slight possibility of a difference in phase shift. This would indicate that during geomagnetically quiet days the energy distribution of the primaries is another in the N direction than in the other two directions, which could be due to the magnetic field of the earth only. From Brunberg's curves we gather that the angle Φ_N is roughly 25° for the N direction and 33° and 35° for the Z and S directions respectively. This is too small a difference in angle to result in such an effect, especially as the geometry of the telescopes was such as to result in a considerable overlapping of the N and Z directions.

We will not enter into further discussion here, as we feel that more measurements are necessary for a thorough understanding of these problems. Also, as far as this kind of measurements is concerned, the investigations should be made with telescopes having comparatively small and well defined apertures.

The dependence of the diurnal variation on the geomagnetic activity raises an interesting point concerning the secular phase shift (Fig. 1). It can be argued that this shift could be due to an increasing frequency of geomagnetic disturbances. If this were true, however, there should not be any or only a small shift in each separate geomagnetic class of days. We find from Fig. 5 that the low resolution leaves us classes II and III only for a study. For better judgement we reproduce the corresponding points in the harmonic dials of Fig. 9.

Although the statistical fluctuations are comparatively great, most of the dials in Fig. 9 put the existence of a secular phase shift even for days with equal geomagnetic activity beyond any doubt. Advancing one step further we correct for the deviation of the primaries in the magnetic field of the earth and plot the points in the diagram of Fig. 10. As in the case illustrated in Fig. 8 we ought to find the same time of maximum regardless of direction. Therefore we are justified in reckoning with the averages of the three values for each date. The existence of a phase shift is quite apparent in both the illustrated cases. It has an average value, indicated by the dotted lines, of 2 hours a year, which is the same as the phase shift for all days regardless of geomagnetic conditions (Fig. 1).

As a result of this part of our discussion we find that the secular phase shift is independent of the geomagnetic activity characterized by the K_p-index. From this it does not follow that it is independent of all geomagnetic effects, as the K_p-index is related only to the daily disturbances, the secular changes being eliminated.

10. Conclusions

1. The secular phase shift of the diurnal variation as found by Elliot and Thambyaphillai is the same regardless of the direction of the telescopes. Accordingly this phase shift depends only on the initial direction of the primaries.

2. The secular phase shift is independent of the daily geomagnetic disturbances as characterized by the K_p-index.

3. The time of maximum amplitude decreases with increasing geomagnetic activity as already found by Sekido and Kodama. Probably, the decrease is smaller than reported by them.

4. The amplitude of the first harmonic increases with the geomagnetic activity.

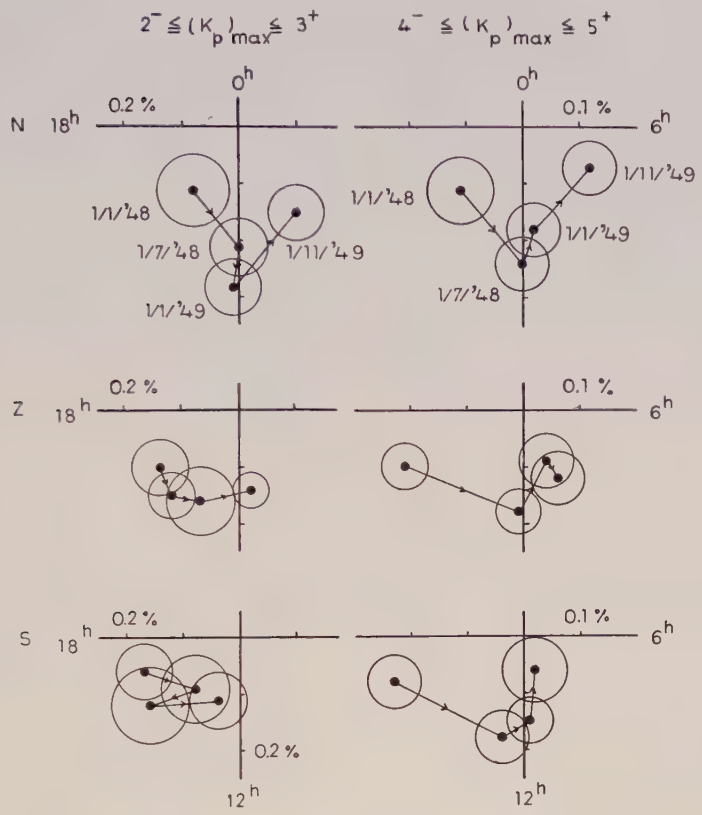


Fig. 9. Harmonic dials for the N, Z, and S directions showing the secular shift for days of equal geomagnetic activity. The dates ascribed to the four consecutive points are the centres of the corresponding intervals as given in Table 1 and Fig. 5.

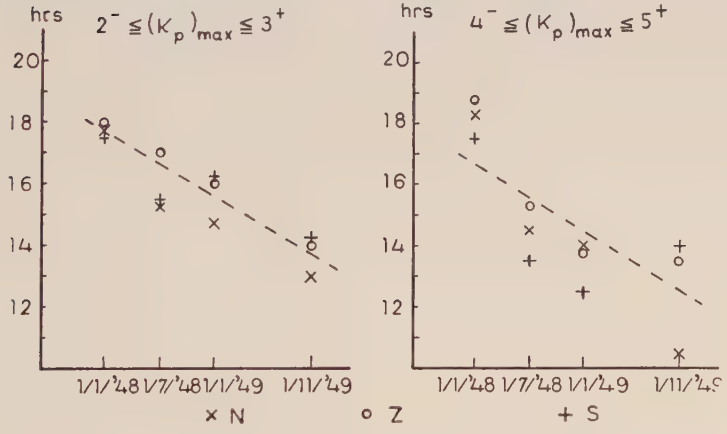


Fig. 10. The secular phase shift for days with equal geomagnetic activity. Corrections have been applied for the influence of the earth's magnetic field. The dates ascribed to the four consecutive points are the centres of the corresponding intervals as given in Table 1 and Fig. 5.

5. When $K_p \geqq 3^+$ the phase shift is certainly independent of the direction of the telescopes, thus indicating a change of the initial direction of the primaries. A discussion shows that this is probably also the case for $K_p < 3^+$.

6. The dependence of the time of maximum on the geomagnetic activity is apparent even as regards averages calculated over periods with an appreciable secular phase shift.

7. Comparing 1, 2, and 5 we conclude that two separate effects exist, both resulting in a change of the initial direction of the primary cosmic ray particles.

11. Acknowledgements

Cosmic ray work in Sweden is being sponsored by the Swedish Natural Science Research Council. The present paper is part of the work carried out under its auspices and directed by its Committee for Cosmic Ray Studies. The author thanks his fellow-members on this comittee for encouragement and discussions, especially Dr K.-G. Malmfors, who made all his measurements available for the present study.

The author also wishes to thank Miss Ulla Björkegren, who made all the numerical calculations.

REFERENCES

Alfvén, H., 1954: On the Origin of Cosmic Radiation. *Tellus.* **6**, p. 232.

Bartels, J., 1949: The standardized index, K_s, and the planetary index, K_p. *International Union of Geodesy and Geophysics*, Association of Terrestrial Magnetism and Electricity. Bulletin No. 12 b, p. 97.

— and Veldkamp, J., 1950: Geomagnetic Indices, *K* and *C*, 1949. *International Union of Geodesy and Geophysics. Association of Terrestrial Magnetism and Electricity.* Bulletin No. 12 c, p. 102—103.

— 1951: Geomagnetic Indices, *K* and *C*, 1950. *International Union of Geodesy and Geophysics. Association of Terrestrial Magnetism and Electricity.* Bulletin No. 12 e, p. 104.

Brunberg, E. Å., 1953: Experimental Determination of Electron Orbits in the Field of a Magnetic Dipole. *Tellus.* **5**, p. 135.

— 1953: On the Anisotropy of the Primary Intensity Time Variations. Recueil des Travaux de l'Observatoire du Pic-du-Midi. Série Rayons Cosmique No 1. Congrès International sur le Rayonnement Cosmique 1953. Session annexe sur les variations du rayonnement cosmique.

— and Dattner, A., 1954: On the Interpretation of the Diurnal Variation of Cosmic Rays. *Tellus.* **6**, p. 73.

Elliot, H. and Thambyaphillai, T., 1953: World-wide changes in the phase of the cosmic-ray solar daily variation. *Nature*, May 23.

Howe, H. H., and Weisman, E. K., 1949: Geomagnetic Indices, *K* and *C*, 1948. *International Union of Geodesy and Geophysics. Association of Terrestrial Magnetism and Electricity.* Bulletin No. 12 b, p. 118—120.

Malmfors, K. G., 1949: Recordings of Cosmic Radiation. *Tellus.* **1**, p. 54.

Sekido, Y., and Kodama, M., 1952: Diurnal variation of vertical cosmic rays. *Rep. of Ionosph. Research in Japan.* Vol. VI No. 2.

Restriction of the Underflow in a Transition

By B. KULLENBERG

Oceanografiska institutet, Göteborg

(Manuscript received August 16, 1954)

Abstract

Bernoulli's Theorem is applied to the problem of the control of salinity in an estuary by a transition.

STOMMEL and FARMER (1953) have developed a theoretical reason to explain why the mouth of a vertically stratified estuary should act as a check on the amount of salt water available for mixing in the estuary. The criterion for "overmixing" was obtained from the equation for a stationary interfacial wave, the equation of continuity, and the equation for the conservation of salt. However, the physical reason why the inflow of salt water cannot rise beyond a certain limit, in spite of a very thorough mixing, is, that an increased inflow of salt water will cause a decrease of the density difference between the ocean and the estuary and, consequently, a decrease of the forces driving the double flow. We shall therefore study the double flow through a transition by applying Bernoulli's Theorem.

We can consider an estuary connecting a river to the ocean through a narrow transition. We assume the mixing to be complete in the estuary as well as in the ocean, at least down to the level occupied by the bottom of the transition, which implies that neither the estuary nor the ocean are vertically stratified above the level referred to. Density and salinity in the estuary are denoted ϱ_1 and s_1, in the ocean ϱ_2 and s_2. The depths of the upper and lower layers in the transition are respectively γ and $D-\gamma$, D being the depth of the transition. The width of the transition, $B = B(z)$, is a function of the depth below the surface. We consider no other forces in the transition than the horizontal pressure gradients due to the higher level of the free surface in the estuary and the density difference between the ocean and the estuary. When ϱ_1 and ϱ_2 occur as factors, we make them equal to one.

In the transition the current velocity as well as the pressure gradient are zero in the interface. At the level z we have a pressure difference $g\,(\varrho_2 - \varrho_1)\,(\gamma - z)$ between both ends of the transition. The flows, Q_1 and Q_2, in the upper and lower layers are

$$Q_1 = \int_0^{\gamma} \sqrt{2\,g\,(\varrho_2 - \varrho_1)}\, B(z)\,(\gamma - z)^{1/2}\, dz \quad (1)$$

$$Q_2 = \int_{\gamma}^{D} \sqrt{2\,g\,(\varrho_2 - \varrho_1)}\, B(z)\,(z - \gamma)^{1/2}\, dz \quad (2)$$

If the discharge, Q_0, of the river exceeds a certain value, Q, there is no double flow in the transition. We find Q by making $\gamma = D$ in (1).

$$Q=\int_0^D \sqrt{2\,g\,(\varrho_2-\varrho_0)}\,B(z)\,(D-z)^{1/2}\,dz \quad (3)$$

where ϱ_0 denotes the density of the fresh water. We have, further

$$Q_1-Q_2=Q_0 \quad (4)$$

$$Q_1 s_1-Q_2 s_2=0 \quad (5)$$

Putting $\varrho_2-\varrho_1=a\,(s_2-s_1)$ and $\varrho_2-\varrho_0=as_2$ we find $\varrho_2-\varrho_1=as_2\,Q_0:Q_1$ and

$$Q=\int_0^D \sqrt{2\,gas_2}\,B(z)\,(D-z)^{1/2}\,dz=A \quad (6)$$

$$Q_1^{3/2}\,Q_0^{-1/2}=\int_0^\gamma \sqrt{2\,gas_2}\,B(z)\,(\gamma-z)^{1/2}\,dz=A_1 \quad (7)$$

$$Q_1^{1/2}\,Q_0^{-1/2}\,(Q_1-Q_0)=$$

$$=\int_\gamma^D \sqrt{2\,gas_2}\,B(z)\,(z-\gamma)^{1/2}\,dz=A_2 \quad (8)$$

Taking the quotients of (7) and (6), and (8) and (6), and denoting by a_1, a_2, q_0, and q_1 the quotients $A_1:A$, $A_2:A$, $Q_0:Q$, and $Q_1:Q$ respectively, we have

$$q_1^{3/2}\,q_0^{-1/2}=a_1 \quad (9)$$

$$q_1^{1/2}\,q_0^{-1/2}\,(q_1-q_0)=a_2 \quad (10)$$

The quantities a_1 and a_2 are determined by the geometric dimensions of the transition, and consequently the inflow of salt water, $q_2=q_1-q_0$, is influenced by the discharge of the river and the geometrical dimensions of the transition but is not dependent on the density of the sea water, when the critical discharge Q is used as a unit. The equations (9) and (10) enable us to determine q_0, q_1, and q_2.

$$q_0=\frac{a_1-a_2}{a_1}\sqrt{a_1\,(a_1-a_2)} \quad (11)$$

$$q_1=\sqrt{a_1\,(a_1-a_2)} \quad (12)$$

$$q_2=\frac{a_2}{a_1}\sqrt{a_1\,(a_1-a_2)} \quad (13)$$

Applying this procedure to *a transition with a rectangular section*, and denoting $\gamma:D=\eta$, we have

$$a_1=\eta^{3/2},\quad a_2=(1-\eta)^{3/2} \quad (14)$$

Elimination of η, using (9) and (10), leads to the equation

$$q_1+q_1^{1/3}\,(q_1-q_0)^{2/3}-q_0^{1/3}=0 \quad (15)$$

The equation is valid for values of q_0 between 0 and 1. Though there is nothing to prevent q_0 from rising above 1, the present problem would not exist in that case, as there is no double flow if $q_0\geq 1$. For a numerical computation it is convenient to use (11), (12), and (13); q_1 and q_2 are given in Fig. 1 as functions of q_0. When q_0 is small we have

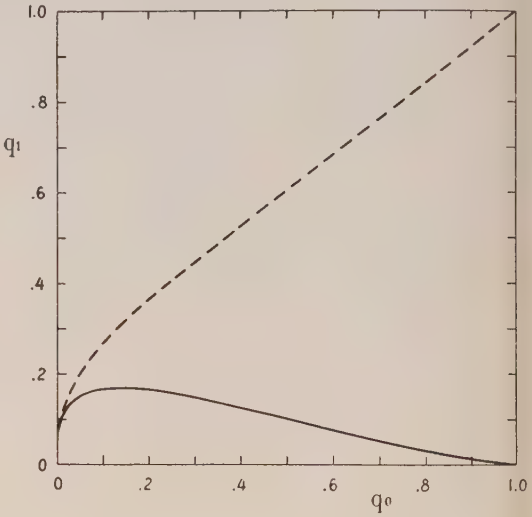


Fig. 1. Upper and lower flow (dotted line resp. full-drawn line) in a transition with a rectangular section, as functions of the discharge of the river. The discharge that is just sufficient to prevent a double flow in the transition is used as a unit.

approximately $a_1=a_2$, or $\eta=1/2$, whence, according to (9), $q_1=1/2\;q_0^{1/3}$. However, as q_0 rises, the relationship very soon becomes practically linear. The flow in the lower layer also rises very quickly with q_0, when q_0 is small. At $q_0=0.142$, q_2 has a maximum $=0.171$, and then q_2 is coming down to zero as q_0 increases to 1. The relationship between η and q_0 is approximately linear, as shown by Fig. 2. With $s_1:s_2=\nu$, then $\nu=q_2:q_1=a_2:a_1$.

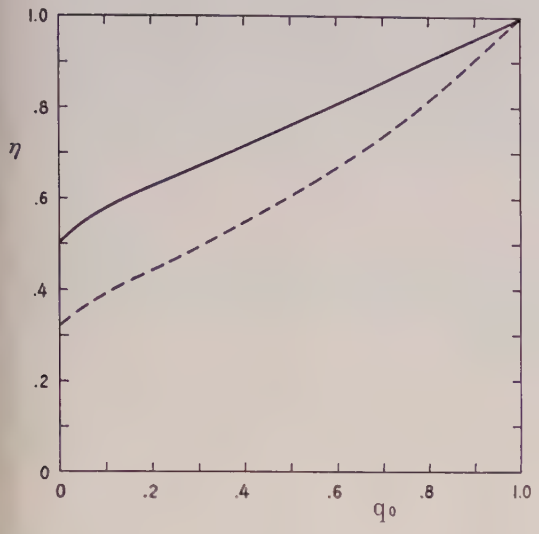


Fig. 2. Ratio of the depth of the upper flow and the total depth of the transition (η) as a function of the discharge of the river. Full-drawn line refers to transition with a rectangular section, dotted line to transition with a triangular section.

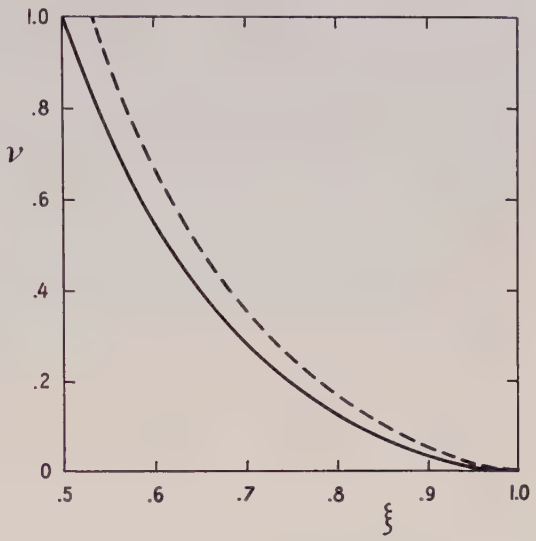


Fig. 3. The ratio (ν) of salinities of the upper and lower layers as a function of the ratio (ξ) of the area occupied by the upper flow and the total area of the transition. Full-drawn line refers to transition with a rectangular section, dotted line to transition with a triangular section.

$$\nu = \left(\frac{1-\eta}{\eta}\right)^{3/2} \tag{16}$$

Applying the same procedure to *a transition with a triangular section*, we have

$$a_1 = \frac{1}{3}(5-2\eta)\,\eta^{3/2} \tag{17}$$

$$a_2 = \frac{2}{3}(1-\eta)^{5/2} \tag{18}$$

It follows that the flow is independent of the shape of the triangle, if the critical discharge Q is used as a unit. The maximal inflow of salt water is smaller than in the case of a rectangular transition, and corresponds to a smaller discharge of the river. By making $a_1 = a_2$ we find that $q_0 = 0$ requires $\eta = 0.316$, which implies that the sectional area occupied by the upper layer is but slightly more than half the sectional area of the transition. Once again the relationship between η and q_0 is approximately linear (Fig. 2). When q_0 is small we have approximately $q_1 = 0.316\, q_0^{1/3}$. The quotient of the salinity in the estuary and that in the ocean is

$$\nu = \frac{2(1-\eta)^{5/2}}{(5-2\eta)\,\eta^{3/2}} \tag{19}$$

In Fig. 3 ν is plotted as a function of the quotient of the sectional area occupied by the upper flow and the total sectional area of the transition. It appears that there is no great difference between the two cases of a rectangular section and a triangular one. It follows that interjacant cases, such as transitions with a trapezium for a section, do not differ appreciably from the two cases treated above.

Any transition with an irregular section can be treated according to (11), (12), and (13).

We have found, among other things, that if the mixing is complete, so as to eliminate vertical stratification, the quotient of the inflow of salt water and the discharge of the river tends to become infinite when the discharge is coming down to zero. Though the case of a less thorough mixing is not treated in this paper, it should be pointed out that the same quotient preserves a finite value, when the discharge is coming down to zero, if the mixing is not complete.

REFERENCES

STOMMEL, HENRY, and FARMER, H. G., 1952: Abrupt change in width in two-layer open channel flow. *J. mar. Res.*, **11**, 205.

— 1953: Control of salinity in an estuary by a transition. *J. mar. Res.*, **12**, 13.

The Particulate Matter in the Sea as Determined by Means of the Tyndall Meter

By N. G. JERLOV, Fishery Board of Sweden, Gothenburg

(Manuscript received December 15, 1954)

Abstract

The Tyndall scattering has been studied for different material occurring in the sea, such as minerogenic matter, calcareous matter, and some marine organisms. It is shown that the scattering is proportional to the total surface of the particles provided that they are larger than 2 μ. This relationship is practically independent of the concentration. The size distribution of particles in marine waters is discussed with a view to estimate the amount of matter removable by sedimentation.

The use of the Tyndall meter offers a simple and quick procedure for current studies of the amount of suspended matter in natural waters (JERLOV, 1953). A parallel Tyndall beam from an electric lamp is projected into the water sample contained in a small glass flask which is placed in a chamber filled with water. The intensity of the light scattered in a forward direction through an angle of 45° is determined by means of a photomultiplier attachment. During the measurement the flask is kept rotating in order to stir up big particles from the bottom of the flask. The particles are maintained in random orientation by the stirring.

The standardization of the method aimed at ascertaining whether the Tyndall reading really does represent the scattering over all angles except a certain minimum angle in the forward direction. The experimental tests furnished sufficient evidence for this; the ratio between the Tyndall reading and the scattering coefficient as derived from measurements with the transparency-meter showed small variations with the particle concentration.

ATKINS and POOLE (1952) recently published an experimental study of the scattering of light by natural waters. Their findings on the angular distribution of the scattered light indicate that the particle scatter through 45° is nearly proportional to that between 20° and 145° for the different water samples (it was not possible to include angles near 0° and 180°). This should support the result of the above-mentioned standardization but it must be observed that much scattering is limited to small angles, less than 20°.

From a physical point of view it appears satisfactory to express the Tyndall reading I_s in terms of a scattering coefficient s. Properly the relation should be

$$e^{-s} = 1 - k\, I_s,$$

where the constant k is found by the standardization. For small values of s the equation is reduced to

$$s = k\, I_s$$

There is a tendency in favour of this latter relationship even for fairly large values of I_s.

Too much emphasis should not be placed on the absolute values of s, which ultimately

depend on the standardization of the transparency-meter (JERLOV, 1953). The minimum angle mentioned is not defined for the transparency-meter and the *s*-values are thus dependent on its construction (cf. GUMPRECHT and SLIEPCEVICH, 1953).

Particle Size

The scattering by particles small compared with the wave-length is given by the Rayleigh equation, the scatter term of which may be written

$$I_s = \text{const. } C\, D^3\, \lambda^4 \qquad (1)$$

where D is the diameter of the particles and C their concentration in the medium.

Rayleigh's Law has been extended to larger particles by MIE (1908) and by JOBST (1925). Their assumption that the particles are homogeneous spheres of known refractive index does not apply exactly to the particulate matter in the sea.[1] It may be inferred, however, that for particles sufficiently large to act as true reflectors, the scattering should be proportional approximately to the square of the diameter. For a dilute system of large particles the scattering may be expressed by the equation

$$I_s = \text{const. } ND^2, \qquad (2)$$

N being the number of particles per unit volume.

As the concentration C is proportional to the volume and so to ND^3, equation (2) can be transformed into

$$I_s = \text{const. } C/D \qquad (3)$$

This approximate relationship has been verified especially by the experiments of TOLMAN *et al.* (1919).

The Rayleigh scattering according to equation (1) occurs for particles of a diameter less than about one-tenth of the wave-length of light. With increasing particle size the power of the diameter decreases from 3 over 0 to −1 until equation (3) is valid. This implies that the scattering for a given concentration passes through a maximum.

[1] Quite recently, BURT (Tellus, *6*, p. 229) has obtained results with the Mie theory which compare favourably with experimental results for the minerogenic suspensions discussed in this paper.

Experiments have been conducted to bridge the gap between equations (1) and (3). STUTZ (1930) has studied suspensions of zinc oxide, and ANDREASEN and his collaborators (1939) have investigated suspensions of solid material in water. As anticipated, these measurements establish the fact that for white suspensions the light scatter has a maximum at grain sizes of the same order as the wave-length of the incident beam.

The significance of the Tyndall reading is realized if equation (2) or (3) is given in the form

$$I_s = K\, S, \qquad (4)$$

where S is the total surface of the particles. The Tyndall reading for large particles thus gives a measure of particle surface (cf. DALLA VALLE, 1948, p. 339). The constant K includes the optical properties of the material (reflectivity, absorption, refraction etc.).

The conclusive results from the experiments mentioned demonstrate that equation (4) is valid down to a size of about 2 μ. Below 2 μ a departure appears which increases toward smaller sizes, and, in addition, the scattering becomes more and more selective.

Particle Colour

The individual particles in a suspension partly scatter, partly reflect, and partly absorb light. These processes exert some selective action on the incident light. The Rayleigh scatter is highly dependent on wave-length [equation (1)]. For large particles the selectivity is chiefly due to their colouring. This effect may be exemplified by some minerogenic suspensions having a high grade of monodispersity (JERLOV and KULLENBERG, 1953). In Table 1 it is shown that the Tyndall reading with red light is 16 % higher than that with blue light for particle sizes 3—12 μ whereas a

Table 1. Ratio of Tyndall reading for red to that for blue. Minerogenic suspensions

Size	I_{red}/I_{blue}
12 μ	1.15
9	1.18
7	1.18
3	1.14
1	0.99

superposed effect of size appears for particles of one μ.

The selectivity due to colour is generally not important. Thus, the existence of brown or red particles in the sea as indicated by a higher scattering for the red than for the blue is a rare phenomenon.

Distribution of Particle Size in the Sea

So far the development concerns uniform particulate matter. The large range of particle size and the variety in particle material in the sea confronts us with new problems. The suspended matter is minerogenic as well as organogenic. The latter component is to a large extent made up of detrital matter, the living fraction being only some percent of the total organic substance (Armstrong and Atkins, 1950; Fox, Isaacs and Corcoran, 1952).

Our knowledge of the distribution of particle size in the sea is rather limited but it is clear that the small sizes predominate in number. This is illustrated by a typical distribution in the surface water found at Bornö Station in the Gullmar Fjord on 1 September 1950 (Table 2). The distribution obtained was a

Table 2. Size distribution of particles in the Gullmar fjord on 1 September 1950

Diameter $D\,\mu$	Number N	Surface ND^2	Volume ND^3
< 2	7,800	8	8
2—4	4,600	40	100
4—10	1,900	90	600
10—20	790	200	3,000
20—40	280	300	8,000
40—100	87	400	30,000
100—200	2	50	7,000
		1,088	48,708

result of microscopic counts performed down to 0.5 μ. In spite of their high number, the contribution of particles below 2 μ to the total surface is negligible.

Goldberg, Baker, and Fox (1952) have studied size distribution of plankton as measured on a molecular filter. From their data it may be derived that sizes of 2 μ contribute less than 1% of the total surface of the plankton. Jenkins and Bowen (1946) made observations on scattering particles in ocean water and found that most of the light cannot be scattered by particles much smaller than the wave-length of the light. Another significant piece of information is furnished by Atkins and Poole (1952). They state: "The measurement with filtered samples show that most of the scattering is caused by particles larger than 1 μ, a small part only being due to particles smaller than 0.2 μ, and relatively little to those of intermediate size."

Similar experiments with filtrations through collodion membranes have been conducted at Bornö Station. The results in Table 3 illustrate

Table 3. Tyndall scattering of unfiltered and filtered samples of surface water at Bornö Station on 20 October 1954

Sample	Tyndall reading
Surface water	840
do, filtered, pore diam. 1.5 μ	52
do, refiltered, p. d. 0.5 μ	60
surface water, filtered, pore diam. 0.5 μ	45

the low percentage of scattering due to particles smaller than 1.5 μ. Filtering must be carried out with utmost care in order to avoid, or to surpress as far as possible the effect of contamination, in particular dust from the vessels and from the air.

It would appear that in the upper strata of the sea the small sizes, below one μ, play a relatively small part in the total surface of the suspended matter. As regards the deep sea the available information on the size distribution is so far quite imperfect but there are several indications that the particles are on the whole smaller than those in the surface layers. From the numerous Tyndall measurements made on deepsea samples the blue scattering by particles comes out 7 % higher than the red, or about the same as for surface samples. This fact alone proves that the selective scattering, which as mentioned begins at a size of 2 μ, is on the whole not significant.

It may be added that some direct observations of the scattering particles in the sea have been made. Emery (1952) found at middepths in the ocean a considerable number of scattering objects as counted from photographic negatives. Most of them were too small for

identification. Similar photographic studies by NISHIZAWA, FUKUDA, and INOUE (1954) revealed the presence of many large objects, mostly above one mm, in coastal waters. They suspect that a disintegration of particles might occur during the process of water sampling.

Correspondence between Tyndall Reading and Particle Distribution in the Sea

Obviously, it is not always possible to include in the Tyndall reading the scattering by occasional large particles. Sometimes a large piece of organic material causes a momentary lighting up in the field of view which effect is not taken into account. Samples from coastal waters often contain relatively large particles to such an extent that an incessant fluctuation appears in the scattering which renders the measurement uncertain, though not impossible. As a rule, the Tyndall reading for oceanic waters gives a satisfactory representation of the average particle content in the flask volume (60 mL) and ultimately in the water bottle (1.2 L) provided that this, when brought onboard, is turned upside down repeatedly, and that the samples are drawn off immediately.

GOLDBERG *et al.* (1952) emphasize that the mass of marine phytoplankton as well as of inorganic suspended material is inhomogeneously distributed in the sea. "The amount of substance collected over a short time interval from a unique location will not necessarily represent conditions in the water region; rather, they may reflect the layering of the water masses or the turbulence of the area."

The discontinuous vertical distribution of particles, which has also been recorded in many transparency measurements, may often find a plausible explanation, say in density variations. Nevertheless it is an indisputable fact as asserted by GOLDBERG *et al.* that irrelevant fluctuations in space and time occur. They suggest that larger samples or integrating collecting devices be utilized to overcome these difficulties.

Considering the Tyndall reading as representative of the particle conditions in the sea the following may be stressed. In studies of the suspended matter its volume or mass is usually concerned. It is obvious that a small number of large particles, even a single large particle, can contribute a considerable part of the total volume (cf. GOLDBERG *et al.*, Table 2). The Tyndall scattering, which records the total surface, is less dependent on size distribution and therefore gives a statistically more satisfactory picture of the particle content in the sea. Thus, the Tyndall method has some advantages which tend to suppress the stated fluctuations in the distribution of the suspended matter.

It is further pointed out that more reliable values are secured by taking average values of readings from various levels in the sea. For instance, the mean particle content between surface and 50 m depth found from Tyndall measurements in the Pacific shows a distribution which is remarkably consistent with the dynamical features (JERLOV, 1953). Another alternative to arrive at a more correct representation would be to repeat the routine water sampling and the Tyndall tests.

Determination of Total Particle Surface

It is of interest to test the validity of the approximate equation (4) for suspensions of different material present in marine waters. This involves a determination of the total surface of the particles which is preferably performed by microscopical studies.

A graded sample of a material is examined under the microscope. Let the number of particles in one litre with the diameter (projected diameter) D_1 be N_1, that of particles with diameter D_2 be N_2 etc. If α_s is the surface shape factor the total surface of the material in one litre will be

$$S = \alpha_s \Sigma N_i D_i^2 \qquad (5)$$

Marine particles of a definite shape, for which the shape factor α_s was readily obtained, or else particle samples for which reliable data of α_s were available, were primarily chosen for the preparation of suspensions. The factor α_s generally keeps within the range 2.0—2.9 (for spherical particles $\alpha_s = \pi$). In the samples the number of particles having a certain projected diameter was determined by direct microscopic measurement and counting after which the surface was evaluated according to equation (5). As a rule, particles down to 0.5 μ were visible in the microscope.

For the minerogenic suspensions mentioned a shape factor of 2.2 was used (cf. HERDAN, 1953, p. 232).

Furthermore, two calcite suspensions were prepared from shell fragments in a sediment core which were selected by sedimentation analysis. These suspensions were to some degree graded, the mean diameter being 30 μ and 10 μ respectively. On account of the high flakiness of the particles a fairly low surface factor, 2.0, was estimated.

The behaviour of the marine organisms as scattering objects was studied in some cases. Organisms were selected with a regular shape preferably spherical so that their surface was readily evaluated. It was aimed at representative selection of the material, calcium carbonate, silica, cellulose etc., with which the species are constructed. Two samples of foraminifera, *viz. Pulleniatina obliquioculata* and *Orbulina universa*, and one sample of radiolaria were obtained by sieving from sediment. The diatom, *Coscinodiscus*, was a living population collected in the sea whereas other suspensions were cultures.

Finally an attempt was made to study a natural suspension, surface water at Bornö Station on 16 June 1953. The constituents were largely minerogenic due to land drainage after heavy rain, but also planctonic and detrital. The large range of sizes rendered the microscopical examination rather laborious. A considerable amount of the minerogenic particles were found to have a size near one μ but their part in the total surface was not important. Some doubt arose as to the plausible shape factor which was eventually estimated at 2.2. In consequence, the evaluated surface value is open to some uncertainty.

Tyndall Measurements

The Tyndall effect was determined for red light (615 mμ) which shows smaller variations with particle material than does the blue light, and which practically does not include any scattering by the water itself.

Most suspensions were gradually diluted with distilled water in order to comprise a scale of different concentrations in the study. The distilled water was produced with due precautions to avoid contamination so that its Tyndall effect was as low as 20. The samples of foraminifera and radiolaria were examined in a single concentration adequately chosen. There was some difficulty in stirring up these big particles even by violent rotation and the readings for them are somewhat uncertain.

The Tyndall effect being read, the counted particles were allowed to settle and the residual scattering due to fine particles or dust was observed. This correction to be deducted from the Tyndall reading was unimportant except for the surface water mentioned. In this case the correspondence between settling time and the lower limit of particles size (0.7 μ) was possibly somewhat misinterpreted.

Relationship between Tyndall Reading and Total Particle Surface

Derived values of total particle surface and Tyndall reading for the uniform minerogenic suspensions are plotted in fig. 1. It is noticed that the curves for sizes 3—12 μ keep together whereas the anticipated departure for the one μsize is considerable. Fig. 1 also shows that the calcite particles produce a relatively more intense Tyndall effect due to their high reflectivity. This is more pronounced for the 30 μ size than for the 10 μ size, the latter particles being to some extent contaminated by a black cover, probably of manganese.

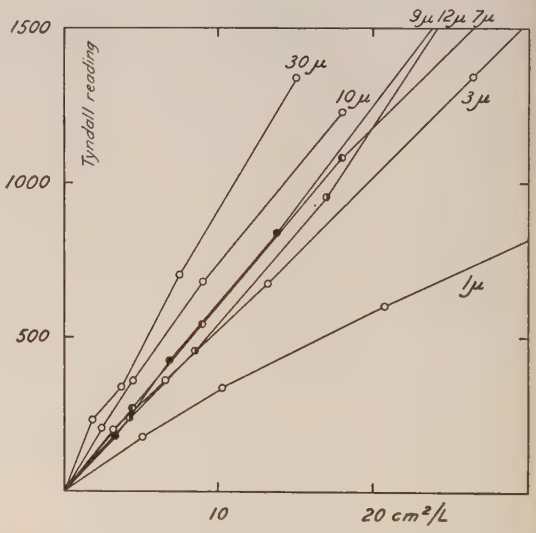


Fig. 1. Tyndall reading and total surface of suspended matter (cm²/L). Uniform minerogenic suspensions with average diameters of 1, 3, 7, 9, and 12 μ, and calcareous suspensions (to the left) with average diameters of 10 and 30 μ.

The entire group of curves clearly demonstrates that the actual relationship is practically independent of the concentration. Thus it is not vitiated by multiple scattering. In other words, the Tyndall reading is proportional to the number of particles per unit volume for the highest concentrations ordinarily attained in the sea.

Table 4. The ratio, *K*, of the Tyndall reading to the total surface of suspended matter

Suspension	*K*
minerogenic 1 μ	32
3	54
7	60
9	59
12	56
calcareous 10 μ	74
30	90
Orbulina universa	79
Pulleniatina obliquioculata	73
Radiolaria	102
Coscinodiscus	98
Chlamydomonas	26
Prorocentrum micans	30
Natural suspension of surface water at Bornö Station on 16 June 1953	80

Table 4 contains the derived values of *K* (equation (4)), also for the biologic samples. The Tyndall effect was relatively high for particles constructed with calcium carbonate or silica whereas the green alga, which besides cellulose contains an abundance of chlorophyll, produced a low scattering. Proportionality between Tyndall reading and number of particles also holds for these biologic samples. The plotted value for the surface water though uncertain fits rather well into the scheme. These characteristics of organisms have an interest of their own, and Tyndall measurements may be useful for instance in studies of plankton cultures.

The question arises whether a mean value of the factor *K* could be chosen so as to obtain from the Tyndall reading the total surface of all suspended matter of the minerogenic material and the organic material, the living as well as the detrital part. This presumes that the water sample shows a normal variety of particles. One must reckon with restricted variations in such a normal distribution but situations with predominance of a certain particle type must be left out.

The large span of the curves in fig. 1 and of the *K*-values in Table 4 does not at first sight seem to lead to a representative average. The deviations from a mean value would be almost $\pm$ 50 %, partly variations due to size and chiefly variations due to particle material.

The presence of fine particles, which produce a lower scattering than equation (4) accounts for, would tend to the total surface becoming underestimated. The above deliberations suggest that fine particles below one μ in the sea contribute little to the total surface though doubt still exists about the size distribution in deep waters. We are thus aware that the surface as directly proportional to the Tyndall reading of a water sample may be underestimated if fine particles dominate. The method must be taken to suffer from this limitation.

As regards the constituents of suspended matter ordinarily occurring in the sea, the preponderance of forward scattering indicates a large amount of rather big, transparent particles (Jerlov, 1951; Atkins and Poole, 1952). Transparent is here taken in the sense that particles reflect and absorb a small part of the incident light. It is manifest that detrital matter is an important component in the sea. This is chiefly made up of resistant remains of organisms such as chitin-, cellulose-, or lignin-like substances (Sverdrup, Johnson and Fleming, 1942) which are not apt to produce a high Tyndall effect. On the other hand, Armstrong and Atkins (1950) found that the ignited residue of suspended matter in the English Channel contained on an average 40 % silica and 20 % calcium carbonate which substances show a higher scattering.

It seems reasonable to conclude that on account of the great variety of particles in an ordinary water sample the Tyndall effect produced by it would be of medium intensity. We may venture an estimate that the Tyndall reading for red multiplied with a constant factor will give the total surface of suspended matter with an accuracy of at least $\pm$ 30 %. In this connection it may be mentioned that the best alternative of determining the surface would be by the laborious measurement and counting with the microscope. Even if microscopic studies are made with great care a

spread in the surface value as high as 20% is obtained.

Polarization of the Scattered Light

We have seen that the Tyndall measurements give little information about the size distribution of particles apart from the fact that they are mainly large. For coastal waters a filtering technique may be resorted to in order to separate size ranges. But this is out of question for deep-sea samples which are easily contaminated. Let us for a moment consider the possibilities offered by polarization measurements.

The light scattered from the Tyndall beam is partly polarized. The degree of polarization varies with 1) optical properties of the particle material, especially the index of birefringence, 2) particle size, 3) concentration of the suspension, 4) angle at which it is observed. HATCH and CHOATE (1930) have studied suspensions of non-uniform particulate matter as silica, granite, and calcite and found the degree of polarization to be a function of the arithmetic mean diameter of the material.

The Tyndall meter with observations through an angle of 45° from the incident beam does not present the most favourable device for polarization measurements as the degree of polarization has its maximum through 90°. Apart from this, it offers no promising chances to apply the polarization method to the mixture of particles in the sea with their large range of size. Only in cases when a certain particle type strongly predominates would it be worth while to employ this procedure. Tests have established that the polarized light as well as the non-polarized must be determined with a high degree of accuracy, otherwise the measurements will be quite useless.

Sedimentation in the Water Samples

By sedimentation of the suspended matter in the flask it would be possible to get an idea of the size distribution. The theoretical background is essentially Stokes' law of settling. GUMPRECHT and SLIEPCEVICH (1953) have made an interesting application of this law. By means of light transmission measurements combined with differential settling they studied particle sizes in polydispersed systems consisting of non-absorbing spherical particles. The method was illustrated by an analysis of an aerosol.

Here we shall only postulate that equation (4), $I_s = KS$, is valid for a suspension of a given material of different sizes. The total surface of the particles per unit volume, S, may be written as

$$S = \pi \int_0^\infty N D^2 \, dD, \tag{6}$$

if N is defined as the number of particles having equivalent diameter D per unit volume of the suspension, per unit range of particle diameter (GUMPRECHT and SLIEPCEVICH).

Thus equation (4) can be expressed in the form

$$I_s = \pi K \int_0^\infty N D^2 \, dD \tag{7}$$

The largest size, D in the Tyndall beam after an elapsed time t is given by Stokes' equation

$$D = \sqrt{\frac{18\, h\mu}{g t\, (\varrho_1 - \varrho_2)}} \tag{8}$$

where h is the settling height during the time t, ϱ_1 and ϱ_2 are the densities of the particle and the liquid respectively, g is the acceleration of gravity, and μ is the dynamic viscosity of the liquid. The settling height, h, is the distance from the upper water surface to the Tyndall beam which is assumed to be small compared with h.

Differentiating equation (7) with respect to t we get

$$dI_s/dt = \pi\, KND^2\, dD/dt \tag{9}$$

into which values of D^2 and dD/dt found from equation (8) are inserted

$$dI_s/dt = -\, K'\, Nt^{-5/2} \tag{10}$$

where

$$K' = \frac{1}{2}\, \pi\, K \left(\frac{18\, h}{g\, (\varrho_1 - \varrho_2)} \right)^{3/2} \tag{11}$$

Using equation (10) the value of N at the time t is found from the observed decrease of the Tyndall reading. The corresponding value of D is obtained from equation (8). The relationship between N and D gives the size-frequency distribution. The above expressions are valid for sizes between 2 μ and 60 μ, the

lower limit being set by equation (4), the upper limit by Stokes' law of settling (equation (8)).

Methods for following the sedimentation have been developed in same detail (see HERDAN, 1953, p. 445). It is particularly important to secure an undisturbed settling in order to avoid thermal convection effects.

This procedure established for material of known density does not apply to the suspended particles in the sea with their large range of density. Nevertheless there is every reason to consider the possibility of reexamining the Tyndall sample after the elapse of a certain time during which the sample is left undisturbed to settle, preferably placed in the Tyndall meter under constant temperature conditions. The time interval must be sufficiently great so that the initial disturbances due to agitation and heating are unimportant, but not so great that any decomposition occurs in the sample. It is proposed that such a time is agreed upon in order to get an idea of the matter removable by sedimentation in water sample. It must be borne in mind, however, that the tranquil conditions necessary for the sedimentation tests do not always exist at sea.

Acknowledgements

The author desires to express his thanks to Dr. K. Fredriksson for much help with the miscroscopic measurements, and to Professor T. Braarud, Dr. H. Höglund and Dr. B. Peters for supplying him with biologic samples.

REFERENCES

ANDREASEN, A. H. M., 1939: The fineness of solids, six lectures at the Royal Danish University, Copenhagen.
Ingeniørvidenskabelige Skrifter, 3.

ARMSTRONG, F. A. J., and ATKINS, W. R. G., 1950: The suspended matter of sea water. *Journ. Marine Biol. Assoc.*, **29,** p. 139.

ATKINS, W. R. G., and POOLE, H. H., 1952: An experimental study of the scattering of light by natural waters. *Proc. Roy. Soc.* B, **140,** p. 321.

DALLAVALLE, J. M., 1948: Micromeritics. New York and London.

EMERY, K. O., 1952: Submarine photography with the bentograph. *Scient. Monthly*, **75,** No. 1.

FOX, D. L., ISAACS, J. D., and CORCORAN, E. F., 1952: Marine leptopel, its recovery, measurement and distribution. *Journ. Marine Res.*, **11,** p. 29.

GOLDBERG, E. D., BAKER, MARJORIE, and FOX, D. L., 1952: Microfiltration in oceanographic research, I. Marine sampling with the molecular filter. *Journ. Marine Res.*, **11,** p. 194.

GUMPRECHT, R. O., and SLIEPCEVICH, C. M., 1953: Measurement of particle sizes in polydispersed systems by means of light transmission measurements combined with differential settling. *Journ. Phys. Chem.*, **57,** p. 95.

HATCH, T., and CHOATE, SARAH P., 1930: Measurement of polarization of the Tyndall beam of aqueous suspension as an aid in determining particle size. *Journ. Franklin Inst.*, **210,** p. 793.

HERDAN, G., 1953: *Small particle statistics.* Amsterdam.

JENKINS, F. A., and BOWEN, I. S., 1946: Transparency of ocean water. *Journ. Optical Soc. America*, **36,** p. 617.

JERLOV, N. G., 1953: Particle distribution in the ocean. *Rep. Swedish Deep-Sea Exp.*, **3,** No. 3.

JERLOV, N. G., and KULLENBERG, B., 1953: The Tyndall effect of uniform minerogenic suspensions. *Tellus*, **5,** p. 306.

JOBST, G., 1925: Zur Farbentheorie kolloidaler Metallsuspensionen. *Ann. Physik*, **76,** 863.

MIE, G., 1908: Beiträge zur Optik trüber Medien, speziell kolloidalen Metallösungen. *Ann. Physik*, **25,** p. 377.

NISHIZAWA, S., FUKUDA, M., and INOUE, N., 1954: Photographic study of suspended matter and plankton in the sea. *Bull. Fac. Fish.*, **5,** p. 36.

STUTZ, G. F. A., 1930: The scattering of light by dielectrics of small particle size. *Journ. Franklin Inst.*, **210,** p. 67.

TOLMAN, R. C., GERKE, R. H., BROOKS, A. P., HERMAN, A. G., MULLIKEN, R. S., and SMYTH, H. D., 1919: Relation between intensity of Tyndall beam and size of particle. *Journ. Am. Chem. Soc.*, **41,** p. 575.

Distribution of Radioactivity in Pelagic Clays[1]

By GUSTAF ARRHENIUS and EDWARD D. GOLDBERG

Scripps Institution of Oceanography, La Jolla, California and California Institute of Technology, Pasadena, California

(Manuscript received September 14, 1954; revised January 29, 1955.)

Abstract

An attempt is made to study the localization of radioelements in different mineral phases of pelagic clay sediments. The results so far obtained indicate that the ionium supported radium is redistributed by diffusion through the interstitial solution and that the authigenic zeolite phillipsite plays an important role as a scavenger of radioelements within the deposit. The mechanism of transfer of radioelements from the ocean to pelagic deposits is discussed, and points of uncertainty in the interpretation of the distribution of gross beta-activity are indicated.

The distribution of radioelements in marine sediments has been extensively studied, primarily for use as a geochronological tool. However, serious shortcomings in dating pelagic sediments have arisen in the cases where uranium and its daughter products have been utilized (a summary of the present situation is given by KRÖLL, 1954). Much of the difficulty in the interpretation of the distribution of uranium, ionium and radium in pelagic sediments stems from a lack of knowledge of the chemical behavior of these elements in marine waters as well as in the sediments.

The purpose of the present investigation was to obtain a deeper insight into the chemistry of the natural radioactive elements in marine deposits through the study of the radioactivity in different components of the pelagic clays.

[1] Contribution of the Scripps Institution of Oceanography, New Series No. 789. This work represents results of research conducted by the University of California under contract with the Office of Naval Research (Contract Nonr-233 (17).

Experimental

The gross beta-activity of clay samples was determined with a windowless flow counter (Nuclear Instrument and Chemical Model

Table 1. Pelagic sediment samples used for the investigation

Sample	Expedition	Core	Latitude	Longitude	Depth	Description
A	Capricorn	33 BG	12° 46′ S	143° 33′ W	4375 m	Mixed material mainly from the lower part of the 128 cm core
B	»	49 BP	9° 17′ N	124° 09′ W	4400	0—24 cm
C	»	21—25 HG + 33 BG	16.5°—17° S	159°—163° W	4375—5150	Mixed material
D	»	31 BG	17° 29′ S	158° 40′ W	4900	4—12 cm
E	»	31 BG	17° 29′ S	158° 40′ W	4900	Mixed material
F	Yo—Yo	7 P	20° 50′ N	125° 11′ W	4520	0—30 cm
	Capricorn	23 HG	16° 59′ S	161° 35′ W	5150	Entire core. Each sample integrating vertically over 4 cm
G	»	39 BG	15° 02′ S	113° 49′ W	2940	Top 4 cm

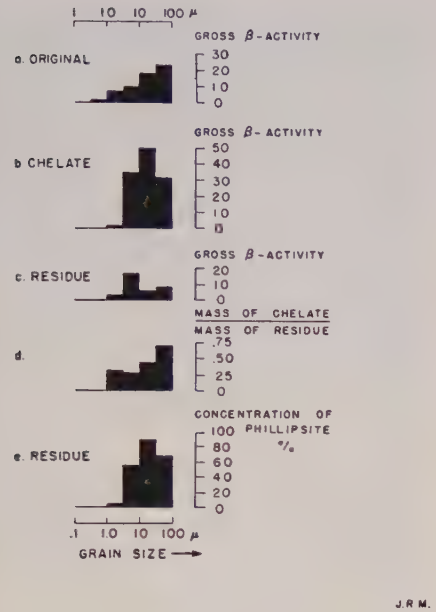


Fig. 1. Distribution of properties in sample A.

D-46 A), in the Geiger region. The powdered sample was compressed into an aluminum planchet with a diameter of 3.1 cm and a depth of 0.2 cm giving 0.1—0.2 g of solid per cm^2 of sample surface. For small samples equally deep planchets were employed with diameters of 1.5 and 1.0 cm and the corresponding net count was multiplied by the appropriate surface factors. The gross beta-activity is mainly due to the decay products of radium and to a minor extent to UX, potassium-40 (about 3 counts per minute), and beta emitters of the thorium family. Loss of emanation would thus markedly influence the results.

The gross beta-activity of the different size fractions of a ferruginous pelagic clay from the South Pacific (Sample A, see Table 1) demonstrated that the coarse fractions are considerably more active han the fine fractions (Fig. 1a and table 2). Since the zeolite phillipsite displayed a similar size distribution (figure 1e), it appeared possible that the activity was associated with this mineral.

To check this assumption, phillipsite in mixed material from the cores *Capricorn 21—25 HG* and *33 BG* was isolated from other minerals by repeated centrifugation of the grains larger than 80 μ in an ethylene tetrabromide-diethyl phthalate mixture with a density of 2.16. The lighter fraction consisted entirely of phillipsite, mostly in the form of single lathes. An intermediate fraction of density 2.16—2.28 contained approximately 95 % of this mineral in the form of crossed twins and spherolithic aggregates, approximately 3 % of manganese micronodules and the rest fossil fragments and unidentified minerals. The heavy fraction with a density greater than 2.50, constituting about 40 % of the size fraction used, consisted mainly of manganese micronodules, with subordinate amounts of feldspar, quartz, phosphatic fossil fragments, fish teeth, foraminiferal tests and 6 % phillipsite. The net activity of the light, intermediate and heavy fractions were 32, 42 and 16 counts per minute respectively.

Radioautographs were prepared of the light (pure phillipsite) and the heavy (phillipsite-poor) fraction by spraying a thin coating of lacquer over a glass slide, distributing the grains over the sticky surface and coating them with another thin spray of lacquer. After drying the specimens, a piece of Eastman No-Screen X-ray film was brought in contact with the mount and exposed for seven weeks. In the light fraction all grains (439) were active enough to give an autograph. In the heavy fraction 130 out of 3,200 grains or 4 % of the sample produced darkening of the film. This is in reasonable agreement with the phillipsite content of 6 %.

Additional autographs were prepared by evaporating suspensions of the light and the heavy fraction of sample D on glass slides, covering the mounts with Kodak liquid

Table 2. β-activity of the size fractions of sample A.

Size Fraction μ	Net β-activity counts/minute	Wt % of total sample	Concentration of phillipsite %
> 32	23.7	6,5	63
10—32	18.9	9.2	89
3.2—10	10.3	12.3	51
1.0—3.2	5.6	22	4
< 1.0	1.8	50	not measured

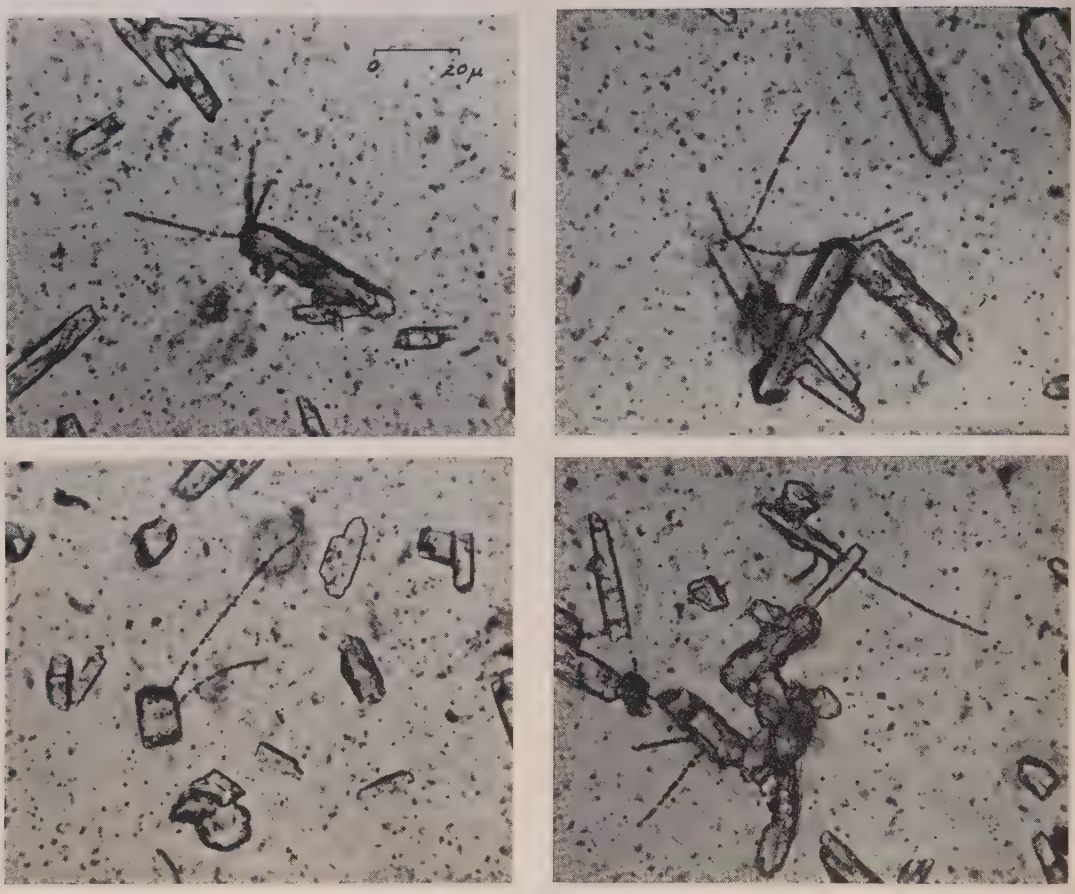


Fig. 2. Radioautographs of alpha particles emitted from phillipsite crystals (sample D) embedded in nuclear track emulsion.

nuclar track emulsion, drying, and exposing for ten days.

Out of 886 phillipsite crystals counted, twenty-five displayed alpha tracks, whereof eight single, nine double, and seven three-branched tracks. In one case eight tracks emerged from one single phillipsite crystal. Microphotographs of some of the active crystals are shown in Fig. 2.

As only the upper side of the mount was coated with emulsion, the recording geometry is approximately 50 %. Most of the single and double tracks observed may therefore be expected to correspond to three-branched stars with one or two of the alpha particles emitted into the glass slide. The high proportion of three-branched stars indicates that the activity is mainly due to an intercept of a disintegration series containing two short-lived alpha emitters. This leaves as the only possibility radium, radon and radium A as the elements responsible. As only in one case a two-branched star was observed in the emulsion without connection to a mineral grain it is probable that the phillipsite retains the emanaftion in an efficient way.

In the heavy (phillipsite-poor fraction) mounted and exposed in the same way as the light fraction, only three mineral grains out of 3,640 counted, displayed alpha tracks, confirming that the radium of the grain size fraction above 5 μ is practically entirely bound to phillipsite.

Most of the activity could be removed by chelation for ½ hour at 100° C. with the neutral ammonium salt of ethylene diamine tetra-acetic acid (Figures 1b – c and Table 3) and by cation exchange using 1N ammonium

Table 3. Chelation of β-active components in sample A.

Size fraction μ	Net β-activity, counts/min	
	Chelate, ashed	Residue
> 32	—	8.2
10—32	30.9	4.7
3.2—10	49.3	14.1
1—3.2	34.4	3.5
< 1	1.5	—

acetate (Table 4). The active constituent occurring in the phillipsite is therefore believed to be sorbed in the pores of the zeolite structure.

The salt from the interstitial solution generally displays a low beta activity which can partly be accounted for by potassium-40 (approx. 3 counts/min). An exception is the interstitial salt from the highly active surface layer of core *Capricorn 31 BG* (Sample D, Table 4 and 5), possibly indicating that in this case a considerable fraction of the radioactive elements are dissolved and not yet in sorption equilibrium with the solid phases.

Sample D displays the highest gross beta activity found in any material so far investigated by the present authors. In contrast to the conditions in the samples A and C, the almost phillipsite-free (4 %) clay fraction displays the same high activity as the pure phillipsite fraction ($> 5\,\mu$, $\varrho < 2.25$) (Table 5).

In order to test the correlation between the stratigraphically abnormal activity and the concentration of phillipsite in the sedimentary sequence, a core was selected (*Capricorn 23 HG*) which displayed in the upper part the type of activity distribution, expected as a result of decay, and in the lower part a marked secondary activity maximum (Fig. 3). All samples from this sequence were analyzed for their phillipsite concentration, using the intensity above background of the X-ray radiation diffracted by planes with the spacing $d_{hkl} = 3.175$ Å. A North American Philips Norelco X-ray diffractometer was used, employing Mn-filtered Fe Kα radiation. The intensities were measured by the time required to accumulate 3,200 counts (40—180 seconds) and the experimental error is approximately ± 3 % of the concentration. The results, shown in Figure 3 indicate that the stratum 44—68 cm, carrying an abnormally high activity, is also characterized by a high content of phillipsite.

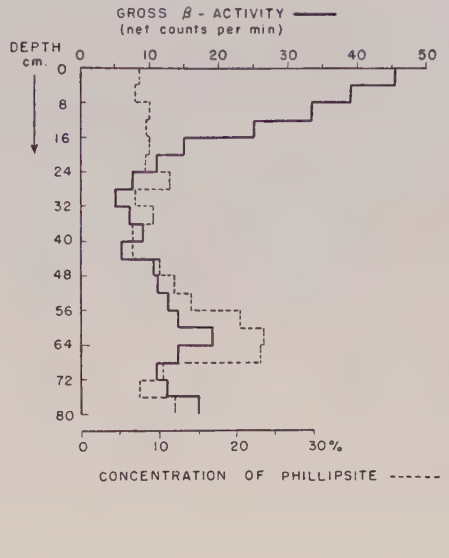


Fig.3. Vertical distribution of gross beta-activity and phillipsite in core *Capricorn 23 HG*.

The above related results can be summarized in the following way: The beta-active con-

Table 4. Net β-activity of salt from interstitial solution, and of sorbed ions exchanged with 1N ammonium acetate.

Sample	Net β-activity counts/min.
B salt.....	4.2
B sorbate.....	67
D salt.....	65
E salt.....	6.5
E sorbate.....	126
F salt.....	9.4
F sorbate.....	34

Table 5. Net β-activity of sample D.

Size fraction μ	Density fraction	Net β-activity counts/min.	Phillipsite conc. %
< 5	—	86	4
> 5	< 2.25	89	100
> 5	> 2.25	36	17

stituents are to a great extent sorbed by phillipsite both in the younger and older strata. The fraction of the deposit which is finer than $5\,\mu$, and which largely consists of clay and hydroxide minerals, is highly beta active in the young sediment, but displays a low activity in the lower parts of the sequence, where significant amounts of unsupported ionium can be expected to be absent. In one case a high beta-activity has been observed in the interstitial salt from the top stratum of the sediment.

The alpha-activity has so far been investigated only in the grain size fraction $> 5\,\mu$ of the young sediment, and is in this to a large extent due to radium and its daughter elements radon and radium A, all bound practically entirely to phillipsite.

Discussion

To explain the distribution of the radioelements observed in marine sediments, the solid matter that accumulated or contained these elements, as well as the chemical reactivity of these elements in the sediments merit consideration (KURBATOV, 1936).

The *lithogenous* (GOLDBERG, 1954) contribution of radioactivity appears generally to be relatively unimportant in the pelagic sediments of the Pacific. This is indicated by the fact (PETTERSSON, 1938; PIGGOT and URRY, 1941) that most of the ionium is not uranium supported.

The accumulation of ionic ionium and radium in hydrogenous oxide-, clay- and zeolite minerals appears to be the quantitatively most important mechanism of transfer of radioelements to the pelagic sediments of the Pacific Ocean. This accumulation could be explained by coprecipitation (scavenging) of the radioelements by the solid phases in statu nascendi, and isomorphous replacement under the same conditions. Scavenging of ionium with ferric hydroxide (PETTERSSON 1953) and with manganese oxides (GOLDBERG 1954) have been suggested. As a considerable fraction of the pelagic clay minerals may be precipitated from dissolved aluminum and silica in the ocean (ARRHENIUS 1954), aluminosilicates may also have to be taken into account as scavengers for radioelements in sea water. The contribution of radioelements by the accumulation of *biogenous* material in the sediments is apparently of less importance than the hydrogenous contribution. EVANS, KIP and MOBERG (1938) found that the shell material in a Pacific mud sample contained but a minor amount of the total radium content of the sediment. Further, core *Capricorn 39 BG*, which has a calcium carbonate content of circa 70 %, shows a beta-activity, based upon its $CaCO_3$-free weight, equivalent to that of the carbonate-free clays in adjacent pelagic areas.

An indication of the minimum contribution of biogenous unsupported radium is given by the radium and barium contents of sea water and pelagic clays, assuming that both elements have similar geochemical paths through the biosphere (*i.e.* fixation in the skeletons and in the soft parts of marine organisms). The barium content of some Pacific pelagic clays attains values of more than 1 % and there is evidence for a biogenous origin of more than 95 % of the barium (GOLDBERG and ARRHENIUS, unpublished). The Ba/Ra ratio in sea water cannot be defined exactly inasmuch as only an upper limit for the Ba-concentration is known. Using the values of $Ra = 3 \cdot 10^{-11}$ mg/l and $Ba < 9 \cdot 10^{-2}$ mg/l, the ratio is $< 3 \cdot 10^{-9}$ and the expected concentration of biogenous, unsupported radium in sediments is $3 \cdot 10^{-12}$ g/g or greater. Since the concentration of Ra in the upper strata of the sediment is between 10 and 70×10^{-12} g/g, the Ba-concentration of sea water needs to be only a few times less than the reported upper limit to make biogenous Ra important as a possible guest element in marine organisms.

Attempts are made at the present time to investigate in detail the distribution of radioelements in stratigraphically surveyed pelagic sediment sequences. On the basis of the present results, and of previous work by HAHOFER and HECHT, KOCZY, KRÖLL, KURBATOV, PETTERSSON, PICCIOTTO and WILGAIN, PIGGOT and URRY, it appears most likely that the high activity of the clay- and hydroxide-mineral fraction at the present sediment surface is largely due to ionium-supported radium.

The abnormal activity of the old strata, originally observed by PETTERSSON (1953), KRÖLL (1954), BARANOV and KUZMINA (1954),

may in some cases be due to displaced radium, sorbed by phillipsite from the interstitial solution. Other possibilities, which have to be investigated, are that the activity maxima in old strata are due to presence of excess uranium (as indicated by some of the results of Hahofer and Hecht, 1954), high concentration of potassium, or to a high retension of radon in the phillipsite, as compared to other minerals present.

The authors are indebted to Dr E. Picciotto for helpful criticism and valuable suggestions.

REFERENCES

Arrhenius, G., 1954: Origin and accumulation of aluminosilicates in the ocean. *Tellus*, **6**, 215—220.

Baranov, V. J., and Kuzmina, L. R. 1954. Ionium method of the age determination of sea sediments. *Doklady Akademii Nauk SSSR* **97**, 483.

Evans, R. D., Kip, A. F., and Moberg, E. G., 1938: The radium and radon content of Pacific Ocean water, life, and sediments. *Am. J. Sci.* **36**, 241—259.

Goldberg, E. D. ,1954: Marine geochemistry 1. Chemical Scavengers of the Sea. *J. Geol.* **62**, 249—265.

Hahofer, E., and Hecht, F., 1954: Determination of uranium in deep sea samples. *Microchimica Acta*, 417—434.

Koczy, F. (unpubl.), quoted in Pettersson, 1945, 178.

Kröll, V., 1954: On the age determination in deep sea sediments by radium measurements. *Deep Sea Research* **1**, 211—215.

Kurbatov, L. M., 1936, Radioactivity of ferromanganese formations in seas and lakes of the USSR. *Nature* **136**, 871.

Pettersson, H., 1938: *Mitt. Inst. Radiumforschung*, 400 a.

— 1945: Oceanbottnens problem. *Ymer*, 161—187.

— 1953: Radium and the deep sea. *Am. Scientist* **41**, 245—255.

Picciotto, E., and Wilgain, S.. 1954: Thorium determination in deep sea sediments. Nature **173**, 632.

Piggot, C. S. and Urry, W. D., 1941: Radioactivity of ocean sediments. III. *Am. J. Sci.* **239**, 81.

A Study of the Time-Dependent Wind-Driven Circulation in a Homogeneous, Rectangular Ocean[1]

By G. VERONIS[2] and G. W. MORGAN[3]

(Manuscript received 29 May, 1954, revised December 10, 1954)

Abstract

A theoretical investigation of the ocean circulation induced by zonal winds whose amplitude varies harmonically with time is presented. The analysis is based on a perturbation scheme with respect to a small parameter which is essentially the frequency of the wind variation and the solutions are derived by boundary layer methods. Results are obtained for the variation of the intensity of the currents and their phase lag behind the wind, as well as for the deviation of the free surface from its equilibrium position.

Introduction

Much of the investigation, both theoretical and observational, in the field of oceanography has centered around the dynamics of ocean currents—including the mass transport of the Gulf Stream and the Kuroshio Current, and the general oceanic circulation. Recently interest has also developed regarding the response of the thermocline (the region of sharp vertical gradient of density) to a time-varying wind.

Since the time of EKMAN's (1905) first paper, a large number of articles dealing with various aspects of ocean currents have appeared in the literature. However, analytical investigations of the problem of general oceanic circulation have met with success only in recent years. In the past decade various interesting and meaningful mathematical models have been suggested by numerous investigators. SVERDRUP (1947), and REID (1948) proposed a fairly simple model which seems to give very good qualitative results for a region with only one north-south boundary. STOMMEL (1948) considered two linearized models with a simplified viscous term. His very important contribution to the overall problem is based on the difference between the results obtained with the two models. In one case, the Coriolis term was constant and the resulting streamline pattern is identical with the one in a model with no rotation. In the second case, the Coriolis term varied linearly with latitude and westward intensification resulted—a factor which was not present in the previous case. Since Stommel's paper all problems dealing with general circulation contain a varying Coriolis parameter. MUNK (1950) refined all the previous work and included the general viscous terms in the equations of motion. He solved the problem of a steady wind blowing over an enclosed ocean, taking account of many of the salient features which are present in the real ocean. Munk's work was extended by MUNK and CARRIER (1950) to include oceans of various geometrical shapes, viz., triangular and semi-circular. It was further extended by MUNK,

[1] The results presented in this paper were obtained in the course of research sponsored by the Office of Naval Research under Contract N7onr-35801, NR-041-032, with Brown University. For a more detailed account of the work the reader is referred to [8] in the bibliography at the end of this paper.

[2] Research Assistant, Division of Applied Mathematics, Brown University, Providence, Rhode Island. Now at the Institute for Advanced Study, Princeton, New Jersey.

[3] Associate Professor, Division of Applied Mathematics, Brown University, Providence, Rhode Island.

Groves, and Carrier (1950) to include the non-linear terms by means of a perturbation procedure.

Along with the American publications, a number of papers have appeared in Japan. Notable among the Japanese authors is Hidaka (1950 a, b, 1951) who published a series of articles covering many of the interesting phenomena of oceanographic problems. Among his contributions are a series of three papers on drift currents in an enclosed ocean and a contribution concerning the neglect of the non-linear terms in the solution of problems in dynamic oceanography (Hidaka, Miyoshi, 1949).

The general problem of oceanic circulation essentially consists of finding the dynamic pattern which results from a given distribution of winds acting on the ocean surface. The complete problem contains a large number of features, such as large-scale oceanic circulation, surface waves, upwelling, etc. To study all these features one would have to take into account the effects of wind, of density and temperature distributions, of the topography of the ocean bed, of the salinity variation, and of many other factors. Needless to say, a mathematical analysis including all these features is impossible. It is therefore necessary to decide what particular aspects of the problem one wishes to study. In this paper we shall confine our attention to the large-scale, wind-driven circulation.

Most of the existing work in this particular field is concerned with steady state motions. In the Atlantic Ocean these include the Gulf Stream and its counter-currents, the Sargasso Sea, etc. It is the purpose of this work to investigate the response of the large-scale motions to a prescribed time variation in the wind. This time-dependent problem has also been considered by Ichiye (1951) and we shall discuss his work later on in this paper.

There is no doubt that the vertical variation of density plays an important role in the response of the ocean to a time varying wind and should therefore be taken into account. For purposes of clarity of exposition, we shall, however, reserve an examination of the effects of this density variation for a future paper and shall assume in the present paper that the ocean is homogeneous, and, moreover, that it is of uniform depth.

Formulation

We assume that the motion which we want to study can be defined mathematically by the Navier-Stokes equations of motion referred to a rotating sphere with the viscous terms replaced by terms arising from a macroscopic viscosity, viz., an eddy viscosity[1]. These equations are

$$\frac{\partial \mathbf{q}}{\partial t} + \mathbf{q} \cdot \nabla \mathbf{q} + \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) + 2\boldsymbol{\Omega} \times \mathbf{q} =$$

$$= -\frac{1}{\varrho} \nabla p + \mathbf{F} + E(\mathbf{q}) \qquad (1)$$

and the equation of conservation of mass is

$$\frac{\partial \varrho}{\partial t} + \nabla \cdot (\varrho \mathbf{q}) = 0 \qquad (2)$$

where $\mathbf{q}$ denotes the relative velocity vector on the rotating sphere,
ϱ denotes the (constant) density,
p denotes the pressure,
$\mathbf{r}$ denotes the radius vector from the center of the sphere,
$\boldsymbol{\Omega}$ denotes the angular velocity vector representing the rotation of the earth,
$\mathbf{F}$ denotes the external forces per unit mass (in our case, gravitation),
$E(\mathbf{q})$ represents the eddy viscosity term which will be defined specifically later.

The full non-linear problem is too complicated to solve analytically. As a result, we shall make a series of simplifying assumptions which we shall list below. The reader is referred to Veronis and Morgan (1953) for a detailed study of some of these assumptions.

1. The fluid is assumed to be incompressible. The continuity equation (2), therefore, reduces to

$$\nabla \cdot \mathbf{q} = 0 \qquad (2\,a)$$

2. The equations on a rotating sphere are approximated by equations in a rectangular

[1] The eddy viscosity term approximately replaces the Reynolds' stresses arising from turbulence. The numerical value of the eddy viscosity coefficient is chosen so as to give a reasonable width to the Gulf Stream.

Cartesian system. The effect of the sphericity of the earth is retained by allowing the Coriolis parameter to depend on the latitude.

3. The vertical acceleration terms and the viscous terms are neglected in the equation of vertical motion so that, in effect, hydrostatic pressure is assumed, i.e., $\partial p/\partial z = -g\varrho$. We let η be the free surface height and require that $p = 0$ at $z = \eta$.

4. The non-linear terms in the equations of motion are neglected. MUNK, GROVES, and CARRIER (1950) have concluded that inclusion of the non-linear terms does not alter the qualitative behavior of the steady state solution. While, in the authors' opinion, this point deserves further careful investigation, we shall assume here that the conclusion of these authors is correct, and moreover, that it also holds in the non-steady case. It should be pointed out, however, that in view of the very drastic simplification made, the purpose of our analysis is the determination of the consequences of stipulating a particular, simple, theoretical model, rather than the investigation of a model which, on the basis of sound reasoning, may confidently be expected to resemble the real ocean.

5. The Coriolis parameter is linearized. In effect, we write $2\Omega \sin\left(\frac{y}{R}\right) \approx \beta y$ where $\beta = 2\Omega/R$. Here, R is the mean radius of the earth, Ω is the angular speed of rotation of the earth and y is the south-north coordinate measured positive northward.

With the above assumptions equations (1) and (2) in component form become

$$\frac{\partial u}{\partial t} - \beta y v = -g\frac{\partial \eta}{\partial x} + A\triangle u + \frac{\partial}{\partial z}\left(A_3\frac{\partial u}{\partial z}\right) \quad (3)$$

$$\frac{\partial v}{\partial t} + \beta y u = -g\frac{\partial \eta}{\partial y} + A\triangle v + \frac{\partial}{\partial z}\left(A_3\frac{\partial v}{\partial z}\right) \quad (4)$$

$$\frac{\partial p}{\partial z} = -g\varrho \quad (5)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \quad (6)$$

where the eddy viscosity terms

$$A\triangle + \frac{\partial}{\partial z}\left(A_3\frac{\partial}{\partial z}\right) \equiv A\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + \frac{\partial}{\partial z}\left(A_3\frac{\partial}{\partial z}\right)$$

are chosen to have the same mathematical form as the viscosity term in laminar flow, subject to the assumption, however, that the kinematic eddy viscosity associated with lateral shear may be different from that associated with vertical shear, the former, A, being taken as a constant, the latter, A_3, as an arbitrary function of position.

The equations (3), (4), and (6) together with the boundary conditions $u = v = w = 0$ on a land-water boundary and an appropriate free surface condition must be solved for u, v, w, η. The wind-stress enters as the value of the shear force at the free surface.

The problem as it has been defined thus far is an extremely difficult one to solve because of the presence of four independent variables x, y, z, t. Consequently, we propose to integrate the equations over the vertical coordinate, z, whereby information regarding the vertical distribution of the velocities is lost. As we are primarily concerned with general oceanic circulation and mass transport, however, and because the integration reduces the complexity of the equations considerably, the advantages gained more than balance the loss of information involved.

Therefore, we choose a constant depth $z = -D$ and integrate the equations from $z = -D$ to the free surface $z = \eta$ (x, y, t). Then the equations (3), (4), become

$$\frac{\partial \overline{U}}{\partial t} - \beta y \overline{V} = -g(D+\eta)\frac{\partial(\eta\varrho)}{\partial x} + A\triangle \overline{U} + A_3\frac{\partial u}{\partial z}\Bigg|_{z=-D}^{z=\eta} \quad (7)$$

$$\frac{\partial \overline{V}}{\partial t} + \beta y \overline{U} = -g(D+\eta)\frac{\partial(\eta\varrho)}{\partial y} + A\triangle \overline{V} + A_3\frac{\partial v}{\partial z}\Bigg|_{z=-D}^{z=\eta} \quad (8)$$

where

$$\overline{U} = \int_{-D}^{\eta} u\varrho\, dz, \qquad \overline{V} = \int_{-D}^{\eta} v\varrho\, dz$$

We have neglected the non-linear terms resulting from interchanging the derivatives and integrals of the velocity terms.

The vertical viscous terms $A_3\frac{\partial u}{\partial z}\Big|_{z=-D}^{z=\eta}$ and $A_3\frac{\partial v}{\partial z}\Big|_{z=-D}^{z=\eta}$ represent the wind-stress

components τ_x, τ_y at the free surface, $z = \eta$, and the bottom friction at the depth, $z = -D$. We assume that the bottom friction terms are negligible and approximate $D+\eta$ by D. Then

$$\frac{\partial \bar{U}}{\partial t} - \beta y \bar{V} = -gD\frac{\partial(\eta\varrho)}{\partial x} + A\triangle \bar{U} + \tau_x \quad (9)$$

$$\frac{\partial \bar{V}}{\partial t} + \beta y \bar{U} = -gD\frac{\partial(\eta\varrho)}{\partial y} + A\triangle \bar{V} + \tau_y \quad (10)$$

If the continuity equation is integrated and the free surface condition (LAMB, 1932)

$$\left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}\right)(\eta - z) = 0$$

is used, the integrated continuity equation becomes

$$\frac{\partial \bar{U}}{\partial x} + \frac{\partial \bar{V}}{\partial y} = -\frac{\partial(\eta\varrho)}{\partial t} \quad (11)$$

The depth $z = -D$ represents either the depth of an ideal ocean of uniform depth or else a constant level above which the mass transports $\bar{U}$, $\bar{V}$ are to be found and at which w is negligible. In either case the analysis assumes that the shear forces at $z = -D$ are negligible.

We now consider a rectangular ocean bounded by land at $x = 0$, $x = r_1$, and by water at $y = 0$, $y = s$. Then the boundary conditions on $x = 0$, $x = r_1$ are that the mass transports $\bar{U}$, $\bar{V}$ vanish there.

The wind-stress is specified to be

$$\tau_x = -(W' + \Gamma' \sin \omega t)\cos ny, \quad \tau_y = 0 \quad (12)$$

where W', Γ' represent the magnitude of the mean wind-stress and the amplitude of the time variation of the wind-stress, respectively,

ω is the frequency of wind variation,

n is the wave number associated with the wind distribution.

One can consider the above form for the wind-stress as a typical term in a Fourier series for a more general wind distribution. The numerical results in this paper are based on a value of ω corresponding to a period of one year, and n is set equal to $2\pi/s$, where s is the north-south length of the ocean, Fig. 1.

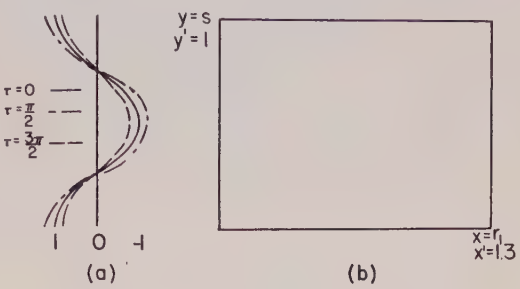


Fig. 1. (a) The wind-stress distribution vs. y coordinate. (b) The rectangular ocean with dimensions in both the dimensional and non dimensional coordinate systems.

The problem to be solved now consists of finding $\bar{U}$, $\bar{V}$, η from equations (9), (10), and (11) together with the boundary conditions $\bar{U} = \bar{V} = 0$ on $x = 0$, r_1, and the specified wind-stress components (12). Since the y derivative of the wind-stress vanishes at $y = 0$, s, plausible boundary conditions at $y = 0, s$ are that these lines be streamlines and that there be no lateral shear, i.e., $\partial\bar{U}/\partial y = 0$.

Solution

The problem as defined in the previous paragraph can be solved by boundary layer methods. In order to apply boundary layer technique it is convenient to non-dimensionalize equations (9), (10), and (11). The choice[1] of the non-dimensional quantities suggested by an inspection of the equations is

$$x' = \frac{x}{s} \quad y' = \frac{y}{s} \quad nW' = W \quad n\Gamma' = \Gamma$$

$$\alpha = \frac{\Gamma'}{W'} = \frac{\Gamma}{W} \quad \tau = \omega t \quad V = \frac{\beta\bar{V}}{W} \quad U = \frac{\beta\bar{U}}{W}$$

$$\varepsilon = \frac{A}{\beta s^3} \quad \delta = \frac{\omega}{\beta s} \quad \Theta = \frac{ngD}{\beta^2 s^3} \quad H = \frac{\varrho\eta\beta^2 s^2}{W}$$

The extent of the rectangular ocean in dimensionless coordinates is given by

$$0 \le x' \le \frac{r_1}{s} \equiv r \quad 0 \le y' \le 1$$

[1] See VERONIS, MORGAN (1953) for details.

Then (9), (10), and (11) become

$$ns\delta \frac{\partial U}{\partial \tau} - nsy\, V + \Theta \frac{\partial H}{\partial x} = ns\varepsilon \triangle U - \\ - (1 + \alpha \sin \tau) \cos nsy \qquad (13)$$

$$ns\delta \frac{\partial V}{\partial \tau} + nsy\, U + \Theta \frac{\partial H}{\partial y} = ns\varepsilon \triangle V \qquad (14)$$

$$\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} = -\delta \frac{\partial H}{\partial \tau} \qquad (15)$$

where we have dropped the primes from x' and y' and will work in the dimensionless coordinate system henceforth.

Attempts to solve equations (13) to (15) in closed form were unsuccessful. We therefore resort to seeking solutions by a perturbation expansion in the parameter δ. This implies "small" δ and hence »small» frequency. We shall learn later just what this means. Let

$$U = U_0 + \delta U_1 + \delta^2 U_2 + \ldots$$
$$V = V_0 + \delta V_1 + \delta^2 V_2 + \ldots$$
$$H = H_0 + \delta H_1 + \delta^2 H_2 + \ldots$$

Our formal procedure is to regard the coefficients U_0, U_1, etc. as coefficients in a power series in δ. If these expansions be substituted into equations (13) to (15), then a series of independent equations will result by equating the coefficients of equal powers of δ on the two sides of the equations.

A. *The zero-order equations.* We have

$$-nsy\, V_0 + \Theta \frac{\partial H_0}{\partial x} = ns\varepsilon \triangle U_0 - \\ -(1 + \alpha \sin \tau) \cos nsy \qquad (16)$$

$$nsy\, U_0 + \Theta \frac{\partial H_0}{\partial y} = ns\varepsilon \triangle V_0 \qquad (17)$$

$$\frac{\partial U_0}{\partial x} + \frac{\partial V_0}{\partial y} = 0 \qquad (18)$$

Eliminating H_0 between (16) and (17) we obtain

$$\varepsilon [V_{0xxx} + V_{0xyy} - U_{0xxy} - U_{0yyy}] - V_0 = \\ = (1 + \alpha \sin \tau) \sin nsy \qquad (19)$$

A stream function ψ may be defined by

$$V_0 = \frac{\partial \psi}{\partial x} \qquad U_0 = -\frac{\partial \psi}{\partial y} \qquad (20)$$

so that (18) is satisfied identically. Then (19) becomes

$$\varepsilon \triangle \triangle \psi - \psi_x = (1 + \alpha \sin \tau) \sin nsy \qquad (21)$$

where $\triangle \triangle \equiv \frac{\partial^4}{\partial x^4} + 2 \frac{\partial^4}{\partial x^2 \partial y^2} + \frac{\partial^4}{\partial y^2}$ is the biharmonic operator and subscripts x, y, τ denote differentiation.

The boundary conditions are

$$\left.\begin{aligned} \psi = \psi_x &= 0 \quad \text{on} \quad x = 0 \quad x = r \\ \psi = \psi_{yy} &= 0 \quad \text{on} \quad y = 0 \quad y = 1 \end{aligned}\right\} \qquad (22)$$

The problem defined by (21) and (22) with $\alpha = 0$ is the one solved by MUNK (1950).

Because ε is small a boundary layer[1] analysis at the boundaries $x = 0$, r is convenient. The solution is found to be

$$\psi = (1 + \alpha \sin \tau) \sin nsy \left\{ -x + r - \varepsilon^{1/3} + \right.$$
$$+ \varepsilon^{1/3} e^{(x-r)\varepsilon^{-1/3}} + \left[(\varepsilon^{1/3} - r) \cos \left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2} \right) + \right.$$
$$\left. \left. + \left(\sqrt{3}\,\varepsilon^{1/3} - \frac{r}{\sqrt{3}} \right) \sin \left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2} \right) \right] e^{-\frac{x\varepsilon^{-1/3}}{2}} \right\} \qquad (23)$$

From (20) U_0 and V_0 are found to be

$$U_0 = -ns(1 + \alpha \sin \tau) \cos nsy \left\{ -x + r - \varepsilon^{1/3} + \right.$$
$$+ \varepsilon^{1/3} e^{(x-r)\varepsilon^{-1/3}} + \left[(\varepsilon^{1/3} - r) \cos \left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2} \right) + \right.$$
$$\left. \left. + \left(\sqrt{3}\,\varepsilon^{1/3} - \frac{r}{\sqrt{3}} \right) \sin \left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2} \right) \right] e^{-\frac{x\varepsilon^{-1/3}}{2}} \right\} \qquad (24)$$

$$V_0 = (1 + \alpha \sin \tau) \sin nsy \left\{ -1 + e^{(x-r)\varepsilon^{-1/3}} + \right.$$
$$\left. + \left[\cos \left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2} \right) + \right. \right.$$

[1] MUNK and CARRIER (1950) used this method for solving the steady problem in a triangular ocean. A detailed discussion of the solution to the problem defined by (21) and (22) is presented in Appendix 5 of VERONIS, MORGAN (1953).

$$+\left(\frac{2r\varepsilon^{-1/3}}{\sqrt{3}}-\sqrt{3}\right)\sin\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right)\Big]\,e^{-\frac{x\varepsilon^{-1/3}}{2}}\Big\} \quad (25)$$

Solving for H_0, we have from (16) and (17) (neglecting terms of order ε),

$$\Theta H_0 = (1+\alpha\sin\tau)(\cos nsy + nsy\sin nsy)\cdot$$
$$\cdot(-x+r-\varepsilon^{1/3})$$
$$+(1+\alpha\sin\tau)\,nsy\sin nsy\Big\{\varepsilon^{1/3}e^{(x-r)\varepsilon^{-1/3}}$$
$$+\Big[(\varepsilon^{1/3}-r)\cos\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right)+$$
$$+\left(\sqrt{3}\,\varepsilon^{1/3}-\frac{r}{\sqrt{3}}\right)\sin\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right)\Big]\,e^{-\frac{x\varepsilon^{-1/3}}{2}}\Big\} \quad (26)$$

B. *The first-order equations.* These are

$$nsU_{0\tau}-nsy\,V_1+\Theta H_{1x}=ns\varepsilon\triangle U_1 \quad (27)$$

$$nsV_{0\tau}+nsy\,U_1+\Theta H_{1y}=ns\varepsilon\triangle V_1 \quad (28)$$

$$U_{1x}+V_{1y}=-H_{0\tau} \quad (29)$$

Eliminating H_1 from (27) and (28), we find

$$\varepsilon\left[V_{1xxx}+V_{1xyy}-U_{1xxy}-U_{1yyy}\right]-V_1=$$
$$=\left[V_{0x}-U_{0y}-\gamma H_0\right]_\tau \quad (30)$$

The boundary conditions are $U_1=V_1=0$ on $x=0$, $x=r$.

Approximate solutions of eq. (29) and (30) may again be found by a boundary layer analysis. The solutions are

$$V_1=\frac{\alpha\cos\tau}{\Theta}(-x+r-\varepsilon^{1/3})\cdot$$
$$\cdot\left[(\gamma^2ns+\Theta n^2s^2)\sin nsy+\gamma\cos nsy\right]+$$
$$+\alpha\cos\tau\sin nsy\,\varepsilon^{-2/3}\Big\{\Big(\frac{1-r\varepsilon^{-1/3}}{3}x+$$
$$+C_2(y)\Big)\cos\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right)+\left(\frac{3-r\varepsilon^{-1/3}}{3\sqrt{3}}x+\right.$$
$$\left.+C_3(y)\right)\sin\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right)\Big\}\,e^{-\frac{x\varepsilon^{-1/3}}{2}}+$$
$$+\alpha\cos\tau\sin nsy\,\varepsilon^{-1/3}\Big\{\frac{x-r}{3}\varepsilon^{-1/3}+$$
$$+A_1(y)\Big\}\,e^{(x-r)\varepsilon^{-1/3}} \quad (31)$$

$$U_1=-\frac{\alpha\cos\tau}{\Theta}\left[2nsy\sin nsy+(\gamma^2n^2s^2+2+\right.$$
$$\left.+\Theta n^3s^3)\cos nsy\right]\left[-\frac{x^2}{2}+x(r-\varepsilon^{1/3})\right]+$$
$$+C_1(y)-\frac{\alpha\cos\tau}{\Theta}nsy\sin nsy\,\varepsilon^{2/3}e^{(x-r)\varepsilon^{-1/3}}-$$
$$-\alpha\cos\tau\,ns\cos nsy\left[A_1(y)-\frac{1}{3}+\right.$$
$$\left.+\frac{(x-r)\varepsilon^{-1/3}}{3}\right]e^{(x-r)\varepsilon^{-1/3}}-$$
$$-\alpha\cos\tau\sin nsy\,\frac{dA_1}{dy}\,e^{(x-r)\varepsilon^{-1/3}}-$$
$$-\frac{\alpha\cos\tau}{\Theta}nsy\sin nsy\cdot$$
$$\cdot\left[(r\varepsilon^{1/3}-2\varepsilon^{2/3})\cos\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right)-\right.$$
$$\left.-\frac{r\varepsilon^{1/3}}{\sqrt{3}}\sin\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right)\right]e^{-\frac{x\varepsilon^{-1/3}}{2}}-$$
$$-\alpha\cos\tau\,ns\cos nsy\varepsilon^{-2/3}\Big[\Big(-\frac{2x\varepsilon^{1/3}}{3}-\frac{\varepsilon^{2/3}}{3}+$$
$$+\frac{rx}{3}\Big)\cos\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right)+$$
$$+\left(\frac{\varepsilon^{2/3}}{\sqrt{3}}-\frac{rx}{3\sqrt{3}}-\frac{2r\varepsilon^{1/3}}{3\sqrt{3}}\right)\sin\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right)\Big]e^{-\frac{x\varepsilon^{-1/3}}{2}}+$$
$$+\alpha\cos\tau\,\frac{\varepsilon^{-1/3}}{2}\cdot$$
$$\cdot\frac{\partial}{\partial y}\Big\{(C_2+\sqrt{3}\,C_3)\sin nsy\cos\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right)+$$
$$+(C_3-\sqrt{3}\,C_2)\sin nsy\sin\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right)\Big\}\,e^{-\frac{x\varepsilon^{-1/3}}{2}} \quad (32)$$

The arbitrary functions of y can be evaluated by means of the boundary conditions $U_1 = V_1 = 0$ on $x = 0, r$. We have

$$\sin nsy\, C_2 = \frac{\varepsilon - r\varepsilon^{2/3}}{\Theta}\left[(nsy^2 + \Theta n^2 s^2)\sin nsy + \right.$$
$$\left. + y\cos nsy\right] \qquad (33)$$

$$\sin nsy\, A_1 = \frac{\varepsilon^{2/3}}{\Theta}\left[(nsy^2 + \Theta n^2 s^2)\sin nsy + \right.$$
$$\left. + y\cos nsy\right] \qquad (34)$$

$$C_1 = \frac{\alpha\cos\tau}{\Theta}\cdot$$

$$\cdot\left\{\left[2nsy\sin nsy + (y^2n^2s^2 + 2 + \Theta n^3 s^3)\cdot\cos nsy\right]\cdot\right.$$
$$\left.\cdot\left[\frac{r^2}{2} - r\varepsilon^{1/3} + \varepsilon^{2/3}\right] - \varepsilon^{2/3}\left(\frac{\Theta ns}{3} + 1\right)\cos nsy\right\} \qquad (35)$$

$$\sin nsy\, C_3 = \frac{1}{\sqrt{3}\,\Theta}\cdot$$

$$\cdot\left\{\left[2\varepsilon^{1/3}\left(y^2 ns + \frac{2}{ns} + \Theta n^2 s^2\right)\left(r\varepsilon^{1/3} - r^2 - \varepsilon^{2/3}\right) + \right.\right.$$
$$\left. + \frac{2\varepsilon}{3} - \frac{2\Theta\varepsilon^{1/3}}{3}\right]\sin nsy +$$
$$+ \left(5y\cos nsy - \frac{4}{ns}\sin nsy\right) r\varepsilon^{2/3} -$$
$$- \left(9y\cos nsy - \frac{10}{ns}\sin nsy\right)\varepsilon^{2/3} +$$
$$\left. + \left(r\varepsilon^{2/3} - \varepsilon\right)\left(y^2 ns + \Theta n^2 s^2\right)\sin nsy\right\} \qquad (36)$$

The first-order contribution to H can be found from equations (27), (28). It is

$$H_1 = -\frac{ns\,\alpha\cos\tau}{\Theta^2}\cdot$$

$$\cdot\left\{\left[(\Theta ns + y^2)\cos nsy + (y^3 ns + y\Theta n^2 s^2)\sin nsy\right]\cdot\right.$$
$$\cdot\left[\frac{1}{2}\left(x^2 + r^2\right) + \left(\varepsilon^{1/3} - r\right)\left(x + \varepsilon^{1/3}\right)\right] + \frac{\Theta}{ns}\cos nsy +$$
$$\left. + \varepsilon^{2/3}\left(\frac{\Theta ns}{3} + 1\right)\left(\frac{y\sin nsy}{ns} + \frac{\cos nsy}{n^2 s^2}\right)\right\} +$$
$$+ \frac{\alpha\cos\tau}{\Theta} nsy\sin nsy\,\varepsilon^{-2/3}\cdot$$
$$\cdot\left\{\left[\frac{\varepsilon^{2/3}}{\sqrt{3}} - \frac{2r\varepsilon^{1/3}}{3\sqrt{3}} - \frac{rx}{3\sqrt{3}}\right]\sin\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right) + \right.$$
$$+ \left[\left(\frac{r}{3} - \frac{2\varepsilon^{1/3}}{3}\right)x + \frac{\varepsilon^{2/3}}{3}\right]\cos\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right) - \frac{\varepsilon^{1/3}}{2}\cdot$$
$$\cdot\left[\left(C_2 + \sqrt{3}\,C_3\right)\cos\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right) + \right.$$
$$\left.\left. + \left(C_3 - \sqrt{3}\,C_2\right)\sin\left(\frac{x\sqrt{3}\,\varepsilon^{-1/3}}{2}\right)\right]\right\}\cdot$$
$$\cdot\, e^{-\frac{x\varepsilon^{-1/3}}{2}} + \frac{\alpha\cos\tau}{\Theta} nsy\sin nsy\cdot$$
$$\cdot\left\{\frac{1}{3}\left(x - r - \varepsilon^{1/3}\right)\varepsilon^{-1/3} + A_1\right\} e^{(x-r)\varepsilon^{-1/3}} \qquad (37)$$

The terms U_1 and V_1 do not satisfy the boundary conditions $V_1 = \partial U_1/\partial y = 0$ on $y = 0.1$. We must recall that these boundary conditions were chosen rather arbitrarily as being plausible ones for the type of wind distribution specified, and the y dependence of the zero-order solution was accordingly chosen as $\sin nsy$. We cannot expect such a y dependence to satisfy the conditions for each set of equations. The fact that U_1 and V_1 do not satisfy these boundary conditions does not seem to be very serious since we do not really know what conditions are appropriate.

If we next consider the equations resulting from equating the coefficients of δ^2 in equations (13) to (15), we obtain

$$\varepsilon\left[V_{2xxx} + V_{2xyy} - U_{2xxy} - U_{2yyy}\right] - V_2 =$$
$$= \left(V_{1x} - U_{1y} - yH_1\right)_\tau U_{2x} + V_{2y} = -H_{1\tau}$$

In the first-order solution the unknown V_1 is of the same order of magnitude as the largest driving term, $V_{0x\tau}$. In the present case, $V_{1x\tau}$ is of order ε^{-1} in the boundary layer and hence V_2 can be expected to be of order ε^{-1} since the equations are completely analogous. By a similar argument, we can expect V_3 to be of order $\varepsilon^{-4/3}$, V_4 to be of order $\varepsilon^{-5/3}$, etc. If we therefore write out the series

$$V = V_0 + \delta V_1 + \delta^2 V_2 + \delta^3 V_3 + \delta^4 V_4 + \ldots$$

we have (in terms of orders of magnitude) near $x = 0$,

$$V = o(\varepsilon^{-1/3}) + \delta\varepsilon^{-1/3} o(\varepsilon^{-1/3}) + \delta^2\varepsilon^{-2/3} o(\varepsilon^{-1/3}) + \\ + \delta^3\varepsilon^{-1} o(\varepsilon^{-1/3}) + \ldots$$

or factoring out the $o(\varepsilon^{-1/3})$, we have

$$V = o(\varepsilon^{-1/3}) [1 + \delta\varepsilon^{-1/3} + (\delta\varepsilon^{-1/3})^2 + \ldots]$$

The perturbation scheme may be expected to be valid provided $\delta\varepsilon^{-1/3} < 1$. We can expect a fairly good approximation from only the first two terms provided the more stringent condition $\delta\varepsilon^{-1/3} \ll 1$ is imposed. If $\delta\varepsilon^{-1/3} = 1/5$, the error involved in neglecting the third term is no larger than 5 % of the first term.

For yearly variation of the wind, $\delta\varepsilon^{-1/3} \approx 1/6$ Hence we shall keep only the first two terms of the series. It should be noted that α determines the magnitude of the effect of the perturbation but it has no bearing on the validity of the expansion.

Discussion of Results

In order to discuss the above solutions, we shall prescribe numerical values for the constants of the problem. Let

$r_1 = 6.5 \times 10^8$ cm $\quad \beta = 2 \times 10^{-13}$ cm^{-1} sec^{-1}

$s = 5 \times 10^8$ cm $\quad D = 5 \times 10^4$ cm

$A = 5 \times 10^7$ cm^2 sec^{-1} $\quad \omega = 2 \times 10^{-7}$ sec^{-1}

$n = 2\pi/s$ $\quad W' = 0.65$ gm cm^{-1} sec^{-2}

The magnitudes of r_1, s, A, D correspond roughly to the Atlantic Ocean parameters. The value of β is chosen so as to give the best approximation to the Coriolis parameter in the latitude of Cape Hatteras. The equality $n = 2\pi/s$ corresponds roughly to the east–west components of the trades and the westerlies. The value of ω corresponds to yearly frequency of the wind variation, and $W' = 0.65$ gm cm^{-1} sec^{-2} is the value used by Munk (1950) for the wind-stress.

Then the dimensionless constants have the values

$$\delta = \frac{\omega}{\beta s} = 2 \cdot 10^{-3} \qquad ns = 2\pi$$

$$\varepsilon = \frac{A}{\beta s^3} = 2 \cdot 10^{-6} \qquad \Theta = \frac{ngD}{\beta^2 s^3} = 0.123$$

$$r = 1.3$$

Also Γ' has been chosen so that

$$\alpha = 0.2$$

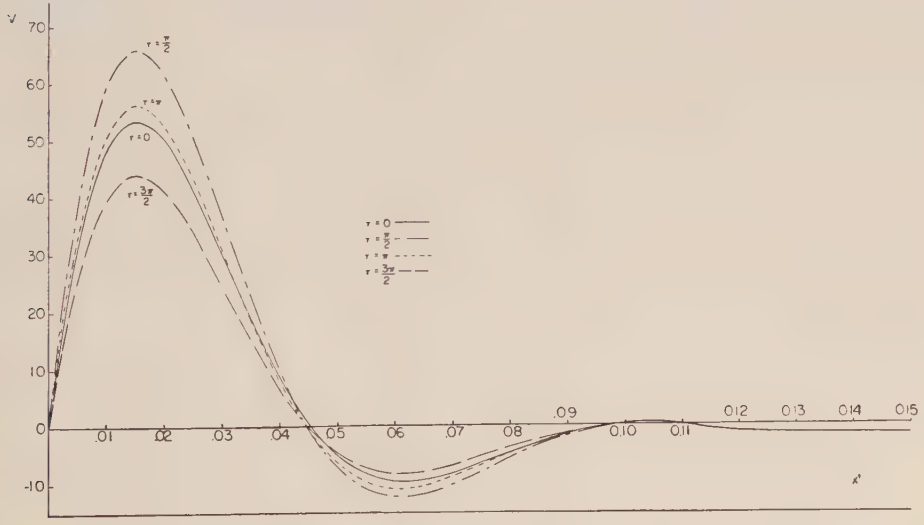


Fig. 2. The non-dimensional mass transport function V near $x' = 0$ (western boundary) showing Gulf Stream and counter current

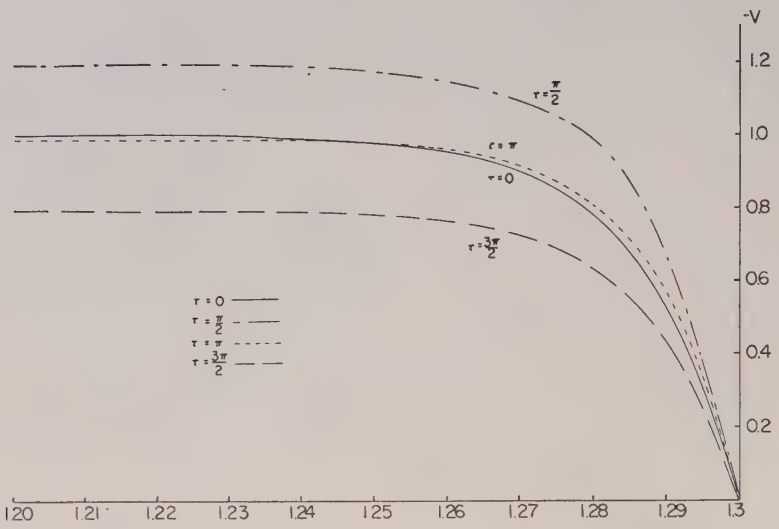


Fig. 3. Non-dimensional mass transport function V near $x' = r$ (eastern boundary)

The results for this numerical example are shown in Figs. 2—6.

In Fig. 2 the non-dimensional, south–north component, V, of the mass transport is plotted against x' near $x' = 0$ for the value $y' = 0.25$. The region of large V corresponds to the Gulf Stream and the section adjacent to the Gulf Stream, with negative V, corresponds to the off-shore counter-current.

Assuming that the complete solution is represented sufficiently accurately by the first two terms of the perturbation series, we see that the transport V is given by $V_0 + \delta V_1$ or $(1 + \alpha \sin \tau)\, Q - \delta\, (\alpha \cos \tau)\, L$ where Q and L are functions of the dimensionless coordinates x and y and of the parameters ε and Θ. Since the wind is given by $1 + \alpha \sin \tau$, this shows that there exists a phase lag, σ, between the wind and the transport such that $\tan \sigma = L\delta/Q$, which, since $\sigma \ll 1$, is approximately given by $\sigma = L\delta/Q$. In terms of dimensional quantities, if $\sigma = \omega t_0$, then, since $\delta = \omega/\beta s$, $t_0 = (L/Q) \cdot (1/\beta s)$ and t_0 gives the lag in the variation of the transport behind that of the wind. The lag is independent of frequency ω and the relative amplitude of the wind variation α. For the transport shown in Fig. 2, the lag is 3 days near the center of the Gulf Stream and $4^1/_2$ days near the center of the counter-current. The maximum transport is $Q(1+\alpha) + \delta^2 \alpha L^2/2Q$ and it occurs at $\tau = \pi/2 + \sigma$. Within the accuracy of the present treatment we can approximate the maximum transport by $Q(1+\alpha)$ since $\delta^2 \alpha L^2/2Q^2 = 0\ (\delta^2 \varepsilon^{-2/3})$, i.e., the correction is of second-order.

It is apparent from Fig. 2 that the out-of-phase effect is of relatively greater importance in the counter-current than in the main stream. The inertial effects introduced by the time dependency give a relatively larger out-of-phase correction and hence, as noted previously, a larger time lag in the counter-current. The graph shows the various effects only up to the eastern edge of the counter-current at $x' = 0.1$. For $x' > 0.1$ only the mean value of the transport is plotted since the deviations from this mean value are very small.

In Fig. 3 the transport V is shown near the eastern boundary of the ocean.

Figures 4, 5, and 6 show surface contours for the southern half of the rectangular ocean for $\tau = 0$, $\pi/2$, π, $3\pi/2$. The contribution of δH_1 is very small throughout the ocean[1] and

[1] If for any of the variables the magnitude of the coefficient of δ in the perturbation solution is of the same order as that of the zero-order term, the coefficient $\delta = 0.002$ renders such a correction negligible. Throughout the present example, the only sizable contribution of the out-of-phase term is found in the north-south transport V in the boundary layer where the function V increases by order $\varepsilon^{-1/3}$. However, H_0 and H_1 have the same order of magnitude throughout the ocean so that the first-order correction H_1 can be neglected throughout.

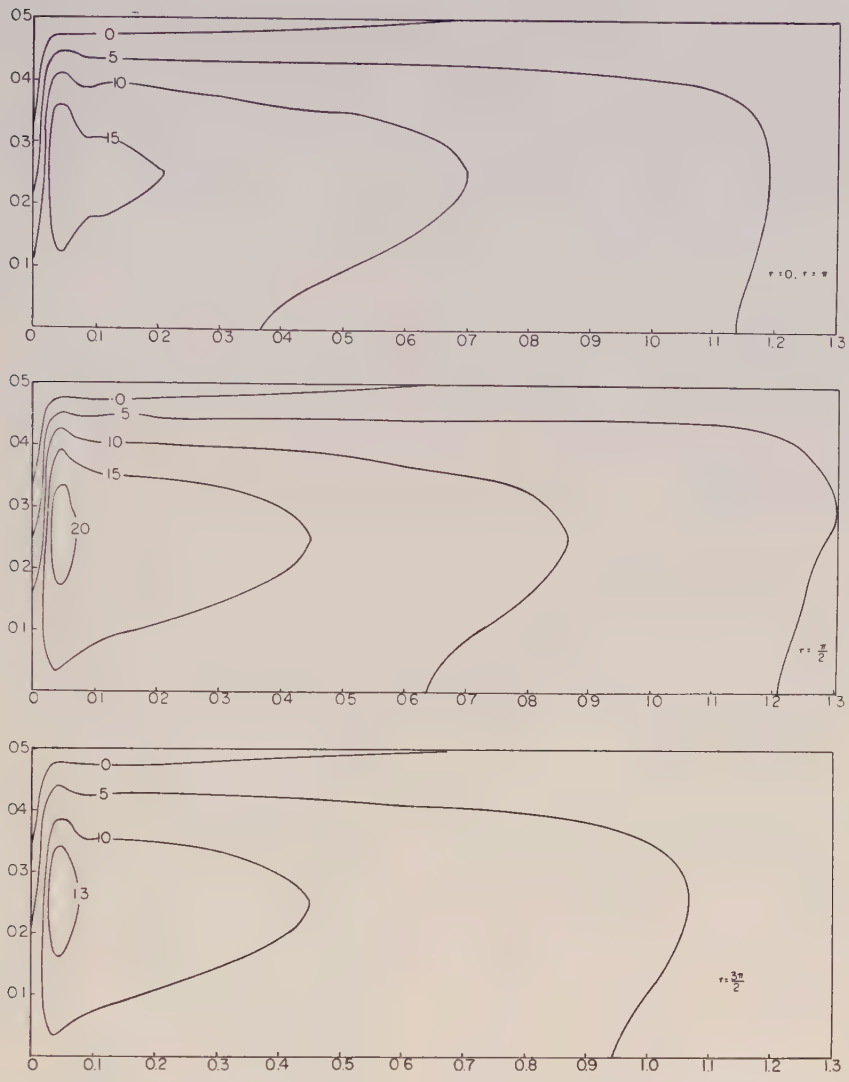


Fig. 4, 5 and 6 — Height in cms of free surface for southern half of rectangular ocean for different »times», $\tau = 0, \pi/2, \pi, 3\pi/2$. The correction of the perturbation terms in negligible.

has therefore been neglected. Thus the graphs for $\tau = 0$ and $\tau = \pi$ coincide.

Our results are based on the assumption that the depth, D, is 500 meters. If we change the value of D, the first-order velocities are also changed since they are dependent on H_0 which is itself directly dependent on D. In particular, had we chosen, say, $D = 250$ meters rather than $D = 500$ meters, the value of H_0 would have been doubled and consequently the values of H_1, V_1, U_1 would have been increased. Their quantitative change must be found by numerical computation.

The above results appear to invalidate the solution of the problem as obtained by Ichiye (1951). Ichiye neglected the contribution of the non-steady term in the integrated continuity equation. However, with the values of the parameters used in this section, the magnitude of this term in the interior of the ocean is as much as ten times that of the remaining non-steady terms which were retained in Ichiye's analysis. Even though we have neglected the contribution of these terms in the interior of the ocean, they are important for the purposes of matching with the boundary layer solution.

The steady mass transport of the Gulf Stream derived here for $\alpha = 0$ is $26.6 \cdot 10^6$ metric tons per second as compared to MUNK's value (1950) of $36 \cdot 10^6$ and the observed value of 72—80 · $\cdot 10^6$ metric tons per second. MUNK (1950) used the east—west component of an empirical wind system and the discrepancy is therefore due to the difference between the two wind systems. The maximum (minimum) transport is $1+\alpha$ $(1-\alpha)$ times the steady transport. In the counter-current the steady mass transport is $4.61 \cdot 10^6$ metric tons per second or approximately 17 % of the mass transport of the Gulf Stream. The relative value of 17 % agrees with observation, though the absolute value $4.61 \cdot 10^6$ is about one-third of the observed counter-current transport.

The difference between the computed and the observed values is not surprising when one considers the many idealizing assumptions made. Such features as the straight coast lines, the simplified theory of turbulence used, the neglect of the non-linear terms, and the idealized stress-effect of the wind on the water could easily combine to yield quantitative results which differ from the observed values by a factor of two or three.

The problem as stated and solved by the above method gives no sensible east–west variation in the position of the Gulf Stream, but a careful investigation of the eastern boundary of the Gulf Stream shows a very small narrowing of the stream. How well such a result agrees with field evidence is uncertain since our solution yields no inshore counter-current.

It would be interesting to ascertain how well our predicted results agree with observation; specifically, if the mass transport of the Gulf Stream responds as indicated to variations in the wind and if the lag of the transport is independent of the frequency.

Acknowledgement

The authors wish to express their appreciation to Mr Henry Stommel for drawing their attention to the problem of non-steady ocean circulation.

Bibliography

EKMAN, V. W., 1905: On the influence of the earth's rotation on ocean currents. *Arkiv för Matematik, Astronomi och Fysik*, **2**.

HIDAKA, K., 1950 a: Drift currents in an enclosed ocean. Part I. *Geophysical Notes (Geophysical Inst. Tokyo Univ.)*, **3**, No. 23.

— 1950 b: Drift currents in an enclosed ocean. Part II. *Geophysical Notes (Geophysical Inst. Tokyo Univ.)*, **3**, No. 38.

— 1951: Drift currents in an enclosed ocean. Part III. *Geophysical Notes (Geophysical Inst. Tokyo Univ.)*, **4**, No. 3.

HIDAKA, K., and MIYOSHI, H., 1949: On the neglect of the inertia terms in dynamic oceanography. *Geophysical Notes (Geophysical Inst. Tokyo Univ.)*, **2**, No. 22.

ICHIYE, T., 1951: On the variation of oceanic circulation, *The Oceanographic Magazine*, **3**, No. 2.

LAMB, H., 1932: *Hydrodynamics*, Dover, 6th edition, pp. 6—7.

MUNK, W. H., 1950: On the wind-driven ocean circulation. *J. Meteor.*, **7**, No. 2.

MUNK, W. H., and CARRIER, G. F., 1950: The wind-driven ocean circulation in ocean basins of various shapes. *Tellus*, **2**.

MUNK, W. H., GROVES, G. W., and CARRIER, G. F., 1950: Note on the dynamics of the Gulf Stream. *J. Marine Res.*, **9**, No. 3.

REID, R. O., 1948: The equatorial currents of the Eastern Pacific as maintained by the stress of the wind. *J. Marine Res.*, **7**, No. 2.

STOMMEL, H., 1948: The westward intensification of wind-driven ocean currents. *Trans. Amer. Geo. Union*, **29**, No. 2.

SVERDRUP, H. U., 1947: Wind-driven currents in a baroclinic ocean; with application to the equatorial currents of the Eastern Pacific. *Proc. Nat. Acad. Sci.*, **33**, No. 11.

VERONIS, G., and MORGAN, G. W., 1953: A study of the time-dependent wind-driven ocean circulation, Office of Naval Research, Report No. A11—101 under Contract N7onr-35801 (Brown University).

Air Borne Salts and the Chemical Composition of River Waters

By ERIK ERIKSSON, Institute of Meteorology, University of Stockholm

(Manuscript received March 14, 1955)

Abstract

The yearly discharge of chloride by rivers in Scandinavia and in the eastern USA is shown on two charts. The isochlor pattern strongly indicates the marine origin of chloride. From the composition of river, lake and ground water compared to rain water it is inferred that the salt in these are to a very great extent air borne.

Some years ago JOEL V. ERIKSSON (1929) published a report on the chemical denudation in Sweden as estimated from the chemical composition of river waters and yearly runoff figures. This investigation was undertaken by the Hydrological Bureau of the Swedish Institute of Meteorology and Hydrology. They collected water samples for chemical analysis at frequent intervals during a number of years at selected points in our Swedish river systems. In the paper by Eriksson the mean chemical composition and the mean yearly runoff are given, as well as the drainage area for each point. It is thus possible to calculate, for each of the constituents investigated, the mean yearly amount carried into the sea from the different areas. Eriksson divided these amounts into a few classes, represented on the charts by different degrees of shading of the areas. Another way of representation is to let the yearly amounts be represented as point values in the drainage area. This has been done here for chloride and the result is shown in fig. 1. The calculations were done in the following way. Firstly, the percentage Cl of the total inorganic content was calculated and from the data on the daily discharge of inorganic compounds, the daily discharge of Cl. (When given, the planemetered data of the total inorganic discharge were used.) The daily discharge of Cl was then multiplied by 365 and divided by the runoff area.

The positions of the points are not always coincident with the centroid of area due to the use of weighting factors dependent on the distribution of precipitation (cf. WALLÉN 1951). On the Swedish west coast an additional weighting factor depending on the distance from the sampling point to the sea has been employed. Only in one case, however, has this latter factor influenced the position, namely, in the case of the Viskan river drainage area where the sampling station was relatively close to the river mouth.

For some areas no measurements of yearly runoff existed at the time the report was published; values for these regions were estimated by Eriksson. Whenever possible these estimates have been checked against more recent data (MELIN 1954) and corrected if necessary. The rivers for which corrections have been made are:

1. *52-1142 and 1143, Himmelsberget Östra and Västra*

Original calculations were based on an average runoff of 13 l sec^{-1} km^{-2}. An adjacent river has a discharge of about 10 l sec^{-1} km^{-2}; this value has therefore been adopted in the present calculations. In addition, there is a misprint in Eriksson's paper (l.c.) with reference to Himmelsberget Västra; the mean amount of inorganic compounds in the runoff is given as 0.041 tons day^{-1}, but a calculation using the given figures for the mean concentration and the runoff estimate yields 0.026 tons day^{-1}.

2. *67-421 Frinnaryd*

The mean runoff is estimated as 9 l sec^{-1} km^{-2}. It is necessary to reduce this figure by 50 per cent since the whole river has a mean runoff of only 6 l sec^{-1} km^{-2}.

Two or more sampling points exist on some

of the rivers. This circumstance makes it possible to subdivide the total drainage area into two or more regions; for example, the chloride amount measured at an upstream station may be subtracted from that determined farther downstream to yield a figure which is representative of the region lying downstream from the first station. This has been done for the Lagan river (98-208 Lagan and 98-221 Knäred 2), river Klarälven (108-274 Edebäck and 108-279 Skåre) and the river Gullspångsälven (108-240 Timsbron and 108-242 Åtorp). In the case of the Nissan river the upstream area (101-1054 Lövrödjan) is so small that no recalculation for the downstream area is necessary. For the river Närkes Svartå only the observation from the station 61-136 Karlslund has been used, as this sampling point represents practically the largest drainage area of this river.

The following rivers are not shown on the map for various reasons:

1. *67-818 Risbro*

River Dummeån, draining into Lake Vättern. The yearly amount of chloride is fairly high, amounting to 38.4 kg ha^{-1} $year^{-1}$.

2. *108-1024 Alstern*

A small lake region draining into the Timsälven-Gullspångsälven river system on the northeast side of Lake Vänern, with chloride drainage of 17.2 kg ha^{-1} $year^{-1}$.

3. *108-1026 Laxbäcken*

A small river draining into the same river system as in (2), with chloride runoff of 24.2 kg ha^{-1} $year^{-1}$.

4. *108-979 Svenningstorp*

The headwaters of river Lidan close to the southwest shore of Lake Vättern. Chloride runoff of 37.2 kg ha^{-1} $year^{-1}$.

These values of chloride runoff agree fairly well with the figures shown on the map of fig. 1, with the exception of 67-818 Risbro.

Bifurcations of rivers occur at two places in northern Sweden. The estimated drainages for each branch as given by Eriksson (l.c.) are used for calculation of the chloride drainage.

Data from Norway and Finland are also shown in fig. 1. The Norwegian data, published by Braadlic (1931) cover a rather important

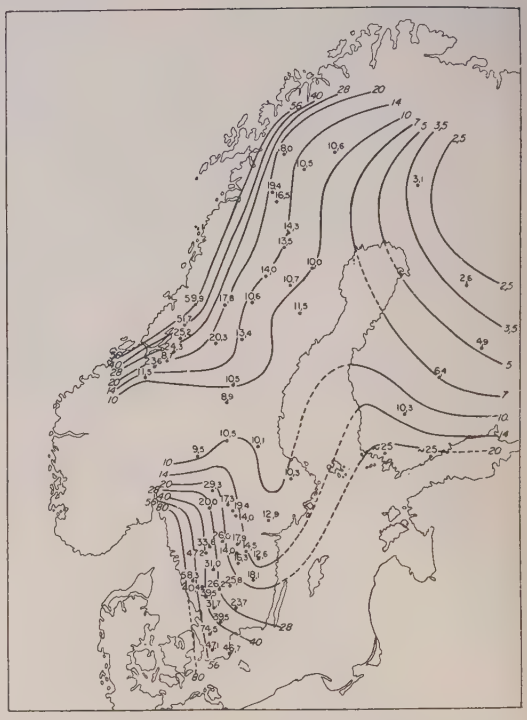


Fig. 1. Discharge of chloride by Scandinavian rivers in kg ha^{-1} $year^{-1}$. Isochlors logarithmically spaced. Note: The figure 74,5 in the southermost part of Sweden should read 52,0.

area and provide valuable information about the west coast of Norway. The Finnish data are obtained from a recent paper by Viro (1953) on the loss of soil nutrients in Finland.

Looking at the plotted values it is immediately seen that there is a general strong decrease in the values (given in kg ha^{-1}) from west to east. For the analysis of this map logarithmically spaced isolines have been chosen, and it is seen that the spacing between the isolines is remarkably constant in some areas. This would be expected if the chloride was airborne in westerly to southwesterly winds and the rate of precipitation of chloride was proportional to the concentration in the air. As a matter of fact the picture in fig. 1 is remarkably similar to that of chloride in rainwater (cf. Emanuelsson et al. 1954 p. 262) which is shown in fig. 2, redrawn from the original data, at least for southern Sweden. The obvious conclusion is that the chloride in river water has mainly been precipitated from the atmosphere. Still, the amounts in rainwater

are appreciably smaller than those in river water, a discrepancy which needs an explanation though some of it may be due to the yearly fluctuations in the Cl precipitation which, as ROSSBY & EGNER (1955) has shown may be considerable.

The contribution of Cl to the river water from the weathering of our bedrocks must certainly be very small. A recent and very interesting survey on the chloride content of bedrocks has been done by BEHNE (1953) in which a number of igneous and sedimentary rocks have been investigated as well as separated minerals. The estimated average content in both types of rocks were very similar, namely 150 and 154 g ton^{-1}. There are rather great variations mostly depending upon the mineral composition. Most of the chloride is chemically bound, especially in apatites, biotites and hornblendes, and is released by the weathering of the minerals. A minor part is occluded in the minerals and this part is easily removed by fine mechanical division. Now, these amounts are very small anyway and cannot possibly contribute anything noticeable. If, for instance, 15 kg ha^{-1} should be taken from the rock about 0.4 cm has to be completely weathered. On an average such a weathering would release about 1,000 kg ha^{-1} of calcium which also would be found in our river waters whereas Eriksson (l.c., 1929) found an average of 10—20 kg in the granite areas. Thus, possibly one or two percent of the Cl in fig. 1 can have come from bedrock.

As parts of Sweden at the end of the last glaciation were submerged, inclusion of chloride into the sedimentary deposits took place and one would expect a certain contribution from this chloride to the river water. On the map in fig. 1 there is, however, no indication of any substantial contribution from earlier submerged areas.

Considering other possible sources, the use of potassium fertilizers in agriculture during the period studied by Eriksson can have furnished on an average over the whole of Sweden 0.5 kg ha^{-1} $year^{-1}$. In agricultural districts like the southernmost province this gives a contribution of about 5 kg ha^{-1} $year^{-1}$ which is still small compared to the actual values. Further north the drainage basins include large areas of uncultivated land and reduces the contribution of fertilizer chloride in the runoff water correspondingly.

An appreciable part of the Cl in river water must be precipitated from the atmosphere in some other way than by rain and snow. Part of the discrepancy can be attributed to the fact that precipitation like rime, dew and fog water are not represented in precipitation analyses and it is known that the concentrations of salts in such forms is very high (cf. ERIKSSON 1952). Further the efficiency of the collecting funnels is certainly low for very fine drizzle, which presumably is rich in Cl. There is, however, also the possibility of direct catch of oceanic salts in small particles by obstacles offered by the vegetation to moving air.

The oceanic salts, from which the chloride is derived, occur in the air as very small hygroscopic particles, so-called condensation nuclei. The settling rate of these is insignificant due to the turbulent character of the air but they can be caught from moving air by small obstacles. Such suitable obstacles are grass leaves, spruce needles and similar small objects. The efficiency of catch is a function of wind velocity and the radius of the obstacle, increasing with increasing wind and decreasing radius. As the small particles are hygroscopic they are generally wet, and once they have been caught they can hardly be removed by the wind. Further, as the air

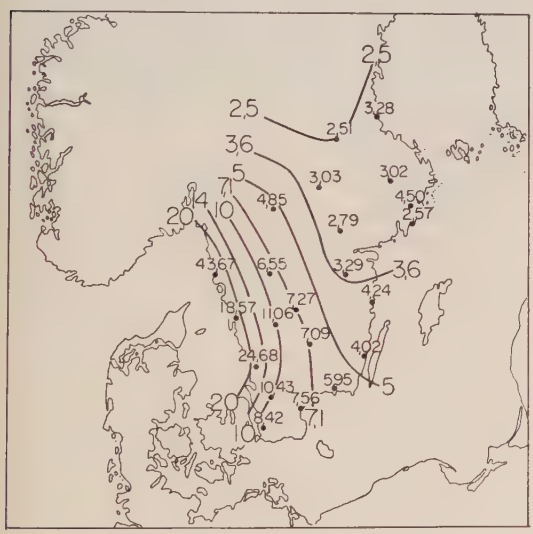


Fig. 2. Chloride in precipitation in southern Sweden in kg ha^{-1} $year^{-1}$. Isochlors logarithmically spaced.

near the ground is depleted, more particles will be transported by turbulent exchange from higher layers; this transport can be very effective especially under strong wind conditions over rough surfaces. A spruce or pine forest is probably the most effective device for this accretion of atmospheric salts which, of course, will be washed down by rain. Tamm (1953 p. 88—90) has recently shown that in eastern central Sweden rainwater collected underneath trees contain much more Na and Ca than when collected in the open (Cl was not determined). In one case after a dry period (l. c. table XX) the concentration of Na in the rainwater collected under the outer parts of a spruce tree crown was five times greater than in the rainwater collected in the open: When the rainwater was collected in the middle of a rain period there was no difference.

The order of magnitude of this nucleus capture by vegetation can be estimated in the following way. Only the "giant" condensation nuclei with radii greater than 1 μ (micron) are large enough to be caught efficiently by spruce and pine needles under the prevailing wind conditions (cf. Woodcock & Gifford 1949, fig. 12). Recent unpublished investigations show that the chloride content of these nuclei in the north of Sweden is of the order of magnitude of one microgram of chlorine per cubic meter of air. If these nuclei are caught by vegetation a density difference of that order would be set up between the "surface" of the vegetation and the moving air. The amount of evaporation of water from vegetation is of the order of some hundred mm per year, i.e. 10^6 kg ha^{-1} year^{-1}. The water vapor density difference between the vegetation "surface" and the air must be of the order of 1 g m^{-3}. Therefore, since the transport mechanism for water vapor is essentially the same (though in the reverse direction), the transport of chloride to the vegetation in the condensation nuclei considered here will be of the order of magnitude of several kg ha^{-1} year^{-1}. This order of magnitude agrees well with the amount of chloride required to give a balance between deposition by rain and observed runoff, so the mechanism appears to be a reasonable one.

The plant ecological importance of such a direct transfer to the ground and vegetation of atmospheric salts need not be stressed.

As to details in fig. 1 it is seen that the south of Sweden, even on high ground, receives a considerable amount of atmospheric salts and that there is a very strong decrease to the north where a minimum area seems to exist. It is, no doubt, the shielding effect of the mountains in South Norway which causes this minimum. Around the 64th parallel there is a broad pass in the mountains allowing a greater salt transport inland, very well indicated in fig. 1. Another interesting feature is that the spacing of the isolines seems to decrease considerably towards the Norwegian west coast. This is probably an elevation effect, the rise of the land from the sea being rather steep compared to the Swedish west coast.

As to Finland it is seen that the northern part is dependent upon the north Atlantic for its chloride while the southern part gets most of its chloride from the Baltic-Sea, though some may have come from more distant parts in southerly winds in which rain is known to contain considerable amounts of chloride (Rossby & Egnér l.c.).

A rather extensive investigation on the chemistry of river waters was carried out in the USA by the US Geological Survey a few years earlier than in Sweden, namely in the period 1905—1908. Water samples from 97 rivers east of the one hundredth meridian were taken frequently during a year's time and analysed for among others Ca^{2+}, Mg^{2+}, $Na^+ + + K^+$, CO_3^{-2}, HCO_3^-, SO_4^{-2}, NO_3^-, Cl^- and total dissolved matter. A report of this investigation was published by Dole in 1909. In another report by Dole & Stabler the same year, yearly data on the runoff from a great number of rivers in the USA are given, together with computed yearly amounts of dissolved solids discharged by these rivers. The computations were to a great extent based on the chemical data published by Dole. From these it is possible to calculate the yearly amounts of Cl in runoff from 64 sampling points in the eastern half of the USA. These computed values have been plotted on the map in fig. 3 approximately in the centre of the drainage basins. For very large river systems greater weight has been given to the lower parts of the basins. Thus for the Mississippi River there are two values, 9.1 and 17.8 west and southwest of Lake Michigan, the first one representing samples collected at

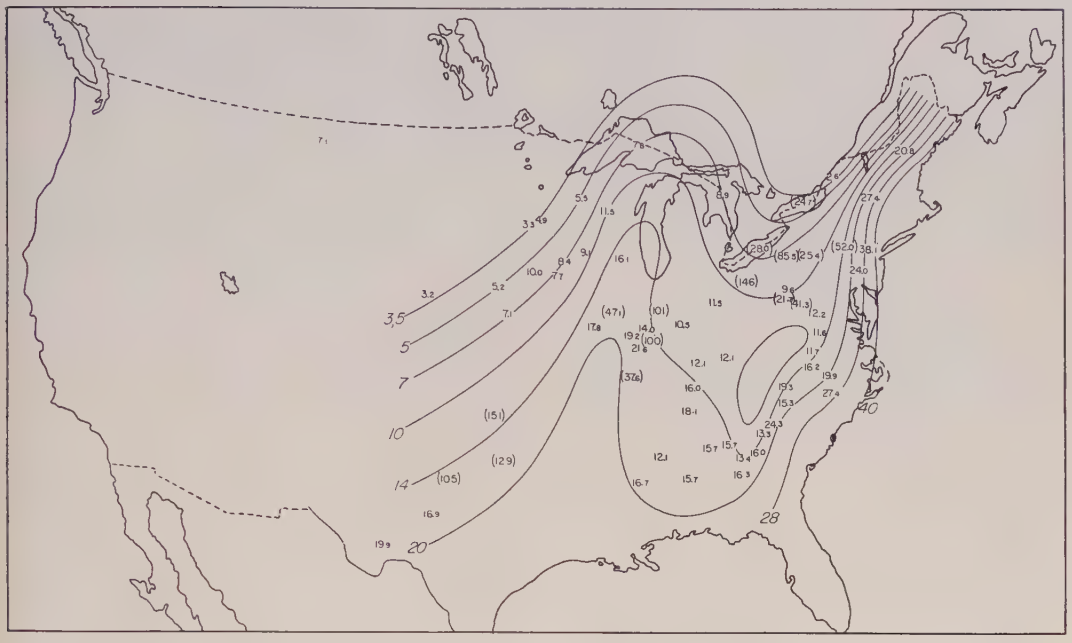


Fig. 3. Discharge of cloride by rivers in the eastern half of the United States in kg ha^{-1} $year^{-1}$. Isochlors logarithmically spaced.

Quincy, Ill., the second samples collected at Memphis, Tenn. The value 19.9 of the Rio Grande represents samples collected at Laredo, Texas.

On plotting the data on the map in fig. 3 it became evident that some values were of a different order of magnitude being up to ten times higher than the bulk of the values. On a closer inspection it appeared that these high values represented oil-producing districts like those of Oklahoma-Texas, Indiana-Illinois and Pennsylvania. It is well known that oil occurs together with salt domes; these high values must therefore be due to contribution from these salt domes. Values which can be suspected to be influenced by such salt deposits are put within brackets on the map and belong to the following rivers and sampling points.

	kg Cl ha^{-1} $year^{-1}$
1. Brazos R., Waco, Tex.	105
2. Arkansas R., Little Rock, Ark.	151
3. Red R. Shreveport, La.	129
4. Illinois R., Kampsville, Ill.	47.1
5. Big Muddy R., Murphysboro, Ill.	37.6
6. Embarras R., Lawrenceville, Ill.	100
7. Wabash R., Vincennes, Ind.	100
8. Muskingum R., Zanesville, Ohio	146
9. Allegheny R., Kittaning, Pa.	85.5
10. Monongahela R., Elisabeth, Pa.	21.7
11. Potomac R., Cumberland, Md.	41.3
12. Susquehanna R., Williamsport, Pa.	25.4
13. " " Danville, Pa.	52.0

It is obvious that Lake Erie receives some terrestrial Cl from such salt domes, hence the value 28.0 for this lake as well as the value 24.7 for Lake Ontario (St. Lawrence R.) have to be discarded. In this way a total of 15 values are neglected and the analysis of the map is based of the remaining 49. Of these one value, 14.0 from Embarras R. at Charlestone has been estimated by the author using about the same figure for total solids in runoff at that place as in the same river at Kampsville.

The isochlors are again logarithmically spaced, and reveal some interesting patterns. As it can be assumed that the Cl is mainly airborne it is seen that there are two main flows of presumably Cl-rich tropical maritime air from the south, one up through the Mississippi Valley and one up along the east

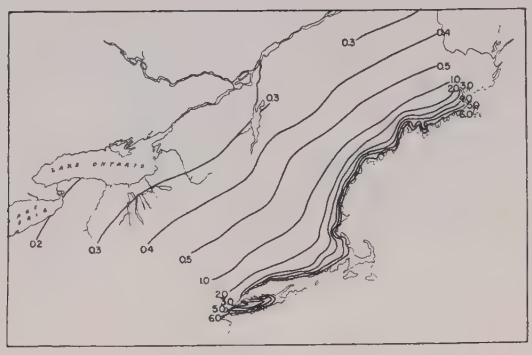


Fig. 4. Chloride concentrations in mg l^{-1} in natural unpolluted waters in New England and New York. (From JACKSON 1905.)

coast, the well known track for Cape Hatteras storms and occasional tropical hurricanes. It is known that tropical maritime air from the Gulf of Mexico can reach farther north than the Canadian border. As there are no topographical obstacles in the broad Mississippi Valley the tropical air can reach very far north before precipitation removes its Cl content.

The low Cl values in the region between the Great Lakes and the east coast can be explained from the fact that most of the precipitation in this region comes from air masses from west which have passed over the mountain ranges in the west and lost its Cl on the western slopes of these mountains. This fact and the frequent passage of Cape Hatteras storm northeastward along the New England coast are responsible for the very steep gradient towards the east coast in this region. The same steep gradient is also seen on the maps of the normal distribution of chlorine in natural waters of New York and New England published by JACKSON (1905). The normal chlorine map over this region published by him is reproduced in fig. 4 and it is seen that the pattern fits nicely into that of fig. 3. According to his isochlors the concentration near the coast amounts to about 6 mg l^{-1} while in the Lake Erie—Lake Ontario region it is only 0.2 to 0.3 mg l^{-1}. For an annual precipitation of 1,000 mm (40 inches) this would contribute 60 kg ha^{-1} $year^{-1}$ near the coast and 2 to 3 kg ha^{-1} $year^{-1}$ at Lake Erie—Lake Ontario.

Jackson's results have been discussed by CONWAY (1942) who used his map to calculate the total amount of sea salts brought annually over land from the sea by multiplying the amount of Cl per unit length of coast line in New England by the length of all coast lines. It is obvious that this will give too low an estimate.

There are very few data available on the content of chloride in precipitation in the USA. In a recent survey (1952) the author has listed three places, namely Geneva, N.Y., Mount Vernon, Iowa, and Madison, Kansas. At Geneva 18 kg ha^{-1} $year^{-1}$ was found as an average for the period 1919—1928 which compared to fig. 3 seems to be somewhat high. COLLISON & MENSCHING (1932) who presented this value admit that pollution by nearby sources was possible as there was a chemical laboratory in the neighbourhood. At Mount Vernon the estimate of Cl in precipitation is 72 kg ha^{-1} $year^{-1}$ which seems to be one order of magnitude too high. In Madison, Kansas, the estimate given is 18 kg ha^{-1} $year^{-1}$, which is in good agreement with the map in fig. 3. From these few data, however, it is not possible to discuss the role of possible direct transfer of Cl from condensation nuclei to the ground.

As to other parts of the North American continent there is very little information available. RAWSON (1942) has published some data on lake waters from the eastern slopes of the Canadian Rockies west of Calgary. They are situated at altitudes between 345 and 1,663 m above sea level and with one exception (the southernmost, Waterton) show only traces of chloride. This can be expected as the air masses reaching these places would have lost their chloride in precipitation earlier.

It is possible from the map in fig. 1 to get some idea about the precipitation of other constituents like Na, K, Mg, Ca, H_3N, NO_3 and SO_4. However, their ratios to chloride in rainwater are not the same as in sea water; generally they are much higher in rainwater, as there may be other sources than the sea (probably distant ones) for these compounds. As to Ca this has been analysed for in the river water, and it may be interesting to compare the Ca/Cl ratio in the river waters with those in rainwaters calculated from the data given by EMANUELSSON et al. (l. c.). In the river waters the Ca/Cl ratio varies from 0.4 to 1.5 with an average around 1 except for river water from limestone areas where the

Ca/Cl ratio may reach values up to 5. In the rainwaters the ratios vary rather systematically from about 1 near the west coast up to 3 or more in the eastern parts. From this it looks as if there were a net gain of Ca in most areas. As the excess Cl in river water compared to rainwater is probably due to the direct transfer of mostly giant condensation nuclei ($> 1\ \mu$) to the ground and as those are mainly composed of Na and Cl (JUNGE 1953) the Ca/Cl ratios in river water and rain water are not comparable. It is possible that the Ca in rainwater represents practically the total Ca delivered to the ground, though there are some indications in TAMM's (l.c.) rainwater analyses that some Ca may be caught by spruce needles. In any case is seems safe to conclude that the denudation of Ca in soil and rock by percolating water is less than that inferred from river water analyses.

As to Na, it can be expected that the Cl/Na ratios in rainwater and river water should be comparable as both occurs in the particles $> 1\ \mu$. It can be mentioned that the Cl/Na ratio in lake waters in Sweden (calculated from the data of LOHAMMAR (1938)) is within the limit of those in rainwater. For Uppland (latitude 60° N, bordering toward the Baltic Sea) for instance, the ratios vary between 0.9 and 1.7 and if some lakes with very low electrolyte content are omitted there is a general trend in the ratios to increase towards the east coast. The same trend is also indicated in rainwater collected at different distances from the Baltic sea in more recent data, yet unpublished. Thus it looks as if Na in lake water is also derived to a very great extent from the atmosphere, and that it even can be relatively less in lake water than in rainwater, suggesting a net uptake by the soil (probably represented by moorland).

The composition of the precipitation when reaching the soil will apparently have a great influence on the composition of ground water as reflected in the river water. The original concentration will be increased not only by dissolved ions from the soil but very substantially by evaporation (including transpiration). As to Cl and probably Na the evaporation effect will be dominant. If the concentration of Cl in the precipitation, when reaching the ground, is C_N, the precipitation N and the evaporation E, then the concentration in the upper parts of the ground water will be

$$C = C_N \frac{N}{N-E} \qquad (1)$$

It is seen that in a region where the evaporation is very near the precipitation very high concentrations can be reached. This may well happen in parts at the flat eastern central part of Sweden and is indicated in Lohammar's analyses from Uppland. GORHAM (1954) found in the analyses by Lohammar from Uppland and Dalecarlia an inverse relation between the total concentration of salts and the height above sea level. He attributed this relationship to the marine submergence of this region during glacial time and the progressive lifting afterwards, more sea salts being washed out in the higher regions than in the lower. As it happens there is also a general increase of precipitation with height above sea level which can be seen from Wallén's map (WALLÉN, 1951), and if the concentration of chloride is plotted against the amount of precipitation as estimated from his map a similar inverse relationship is obtained, showing the effect of evaporation, which at any rate is practically constant in this region compared to the precipitation.

RODHE (1949) made a very interesting study of Lohammar's lake water analyses. He found a remarkable similarity in the percentage composition of these lake waters, though the actual concentration differs considerably. This would be expected from eq. 1 if most of the salts in the lake waters were airborne, the concentration depending upon the $\frac{N}{N-E}$ ratio. This also has other implications, namely that the content of exchangeable ions in the upper parts of soils is to a large extent determined by two factors, the composition of the precipitation as it reaches the ground, and the composition of the organic matter which binds the exchangeable cations. As a steady state presumably has been reached the percentage composition of the precipitation will remain unchanged when percolating through these layers. Deeper down modifications may occur, especially with respect to Ca in our limestone areas, but additions of Cl is normally very unlikely and it looks as if Na too is mainly of atmospheric origin. As to the earlier submerged areas of eastern central Sweden it can be mentioned that MATTSON et al. (1950) investigated soil profiles on

sedimentary clays and found only traces of Cl down to about 6 metres depth where a sudden increase occurred. This depth, however, was three metres down in the so-called soap clay, which is practically impermeable to water. As practically all the drainage occurs in the upper parts, the so-called dry crust, practically no chloride can come from the sedimentary clays.

It is apparent from the discussion above that the concentration of salts in ground water cannot be used as any measure of the chemical denudation neither qualitatively as suggested by many and most recently by Troedsson (1955) nor quantitatively as claimed by Arrhenius (1954). Arrhenius has presented a very interesting material, however, and revealed some very interesting relationships which ultimately may lead to a much better understanding of the hydrology of bedrocks but it looks from his data as if the deep groundwater contributes but little to the yearly water circulation. The river water is most probably derived from the uppermost parts of the ground water.

The best approach made so far towards the estimation of chemical denudation has been made by Viro (l.c.) who subtracted the salts in precipitation from those in river water.

Acknowledgment

I wish to express my thanks to Professor C.-G. Rossby for valuable suggestions and encouraging interest during the preparation of this paper.

REFERENCES

Arrhenius, O., 1954: Chemical denudation in Sweden. *Tellus*, **6**, 326—341.

Behne, W., 1953: Untersuchungen zur Geochemie des Chlor und Brom. *Geochim. & Cosmochim. Acta* **3**. 186—214.

Braadlie, O., 1931: Om elvevannets sammensetning i Trøndelag. *Kgl. Norske Videns. Selsk. Skrifter 1930* nr 5, 52 pp.

Collison, J. B. & Mensching, J. E., 1932: Composition of rain water at Geneva, N.Y., for a 10-years period. *N.Y. State Sta. Techn. Bul.* **193**, 19 pp.

Conway, E. J., 1942: Mean geochemical data in relation to oceanic evolution. *Proc. Royal Irish Acad.* **48**, 119—159.

Dole, R. B., 1909: The quality of surface waters in the United States. Part I. Analyses of waters east of the one hundreth meridian. *U.S. Geol. Survey. Water-Supply Paper* 236.

Dole, R. B. & Stabler, H., 1909: Denudation (Papers on the conservation of water resources). *Ibid.* 234.

Emanuelsson, A., Eriksson, E., & Egnér, H., 1954: Composition of Atmospheric Precipitation in Sweden. *Tellus* **6**. 262—267.

Eriksson, J. V., 1929: La Dénudation Chimique en Suède. *Medd. från Statens Meteorologisk-Hydrografiska Anstalt.* Bd 5 nr 3, 96 pp.

Eriksson, E., 1952: Composition of atmospheric precipitation. Part II. *Tellus* **4**, 280—303.

Gorham, E., 1955: On some factors affecting the chemical composition of Swedish fresh waters. *Geochim. & Cosmochim. Acta* **7**, 129—150.

Jackson, D. D., 1905: The normal distribution of chlorine in natural waters of New York and New England. *U.S. Geol. Survey. Water-Supply and Irrigation Paper* 144, 31 pp.

Junge, Ch., 1953: Die Rolle der Aerosole und degasförmigen Beimengungen der Luft im Spurenr haushalt der Troposphäre. *Tellus* **5**, 1—26.

Lohammar, G., 1938: Wasservegetation und höhere Vegetation Schwedischer Seen. *Symb. Bot. Upsalienses* **3**: 1. 252 pp.

Mattson, S., Eriksson, E., Koutler-Andersson, E., Barkoff, E., Ståhlberg, S. & Vahtras, K., 1950: Phosphate relationships at soil and plant. VII. Forms of *P* in the Marsta hydrologic day series. *Ann. Royal Agricultural Coll. Sweden*, **17**, 222—246.

Melin, R. 1954: The discharge in the Swedish Rivers. *Met. & Hydrol. Inst. of Sweden*, Comm. ser. D. No. 6.

Rawson, D. C., 1942: A comparison of some large alpine lakes in Western Canada. *Ecology* **23**, 143—161.

Rodhe, W., 1949: The Ionic Composition of Lake Waters. *Proc. Intern. Assoc. Limnology* **10**, 377—386.

Rossby, C. G. & Egnér, H., 1955: On the Chloride-Sodium Ratio in Atmospheric Precipitation as a Function of the Prevailing Circulation Pattern. *Tellus* **7**, 118—133.

Tamm, C. O., 1953: Growth, Yield and Nutrition in Carpets of a Forest Moss (Hylocomium Splendens). *Medd. Statens Skogsforskningsinst.* **43**: 1, 140 pp.

Troedsson, T., 1955: Vattnet i skogsmarken. *Ann. Royal School of Forestry* No. 20.

Viro, P. J., 1953: Loss of nutrients and the natural nutrient balance of the soil in Finland. *Communicationes Inst. Forestalis Fenniae* **42**: 1, 51 pp.

Wallén, C. C., 1951: Nederbörden i Sverige. *Medd. Sveriges Met. & Hydrol. Inst.* Ser. A nr. 4.

Woodcock, A. H. & Gifford, M. M., 1949: Sampling atmospheric sea salt nuclei over the ocean. *J. Mar. Research* **8**, 177—197.

The Acidity of Scandinavian Precipitation

By

EARL BARRETT, Institute of Meteorology, University of Stockholm and Woods Hole Oceanographic Institution[1]

GUNNAR BRODIN, Royal Agricultural College, Uppsala

(Manuscript received March 3, 1955.)

Abstract

Data on the pH of the total monthly precipitation at stations of a Swedish network for sampling and chemical analysis of precipitation and atmospheric aerosols during the year July 1953—June 1954 are presented and discussed, together with the pH data from the first two months of operation of a larger pan-Scandinavian net. It is found that well-defined regions of acidity and alkalinity relative to the pH of water in equilibrium with atmospheric carbon dioxide exist, and that these regions persist to such an extent that the monthly deviations from the pattern of the annual mean pH at each station are relatively small. The mean pH of the entire network shows a distinct annual cycle, with greatest acidity in winter and greatest alkalinity in late spring. Coastal stations unaffected by local pollution show persistently high acidity, while inland northern stations show equally persistent alkalinity. Some possible reasons for the observed distributions are considered.

Introduction

Although there is doubtless a considerable number of spot observations of the acid-alkaline balance of precipitation which have been made and reported in various journals during the past, there has been no effort (so far as is known to the writers) to systematize such observations and present them in a significant and readily-assimilable form. One can readily recognize two reasons for this; the first reason is that the observations had been made for the most part by workers who are not trained in the concepts and techniques of synoptic observations and analysis, while the second is that, even if the observers felt the desirability of a synoptic presentation, the necessary organized network of stations for the purpose did not exist. Within the last few years both of the obstacles have been overcome to some degree. A network of stations has been established in Sweden for the purpose of analyzing one-month samples of precipitation for the principal ionic constituents. This net, originally established for agricultural purposes by the Royal Agricultural College of Sweden, has been extended by degrees, until, at the time of writing, it provides excellent coverage of Sweden and moderate coverage of Norway, Denmark and Finland. This network and the sampling equipment have been described in previous papers (Egnér and Eriksson, 1955; Rossby and Egnér, 1955).

The second obstacle has been removed by the establishment of a co-operative program between the Agricultural College and the Meteorological Institute of the University of Stockholm. The junior author of this paper has had the responsibility for the collection and chemical analysis of the samples, while the senior author has undertaken the work of synoptic analysis and meteorological interpretation of the data.

[1] Contribution No. 785 from Woods Cole Oceanographic Institution.

The objective of the paper is to present the results of a preliminary survey of the spatial and temporal variation of the acid-alkaline balance of precipitation. It should be made clear at the outset that this paper will pose many problems and solve few, if any, of them. The atmosphere, considered as a chemical system, is so highly multivariate that it resembles a biological or sociological system in complexity; thus one cannot expect to find the simple logical connections between single causes and effects which one looks for in connection with well-isolated physical systems. The most that one may expect in the preliminary stages of study of such a complex system is to narrow down the field of investigation by looking for the highest correlations, and to subject these to more searching study with the aid, wherever possible, of theoretical models and idealized laboratory experiments. The objective of this paper is the modest one of displaying the data in synoptic form and laying a number of study topics before the attention of others, in the expectation that interest will be stimulated and answers to the questions will be sought by investigators at other institutions at the same time that they are being pursued in Stockholm.

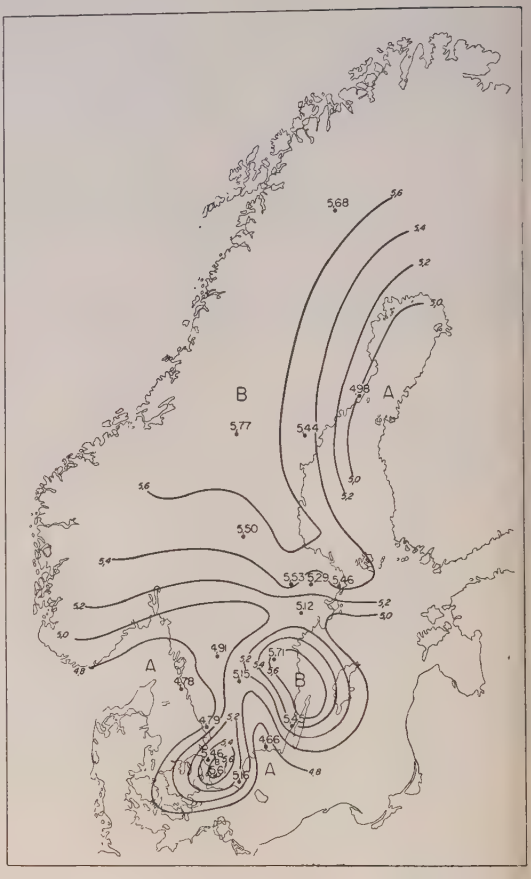


Fig. 1. Annual Mean pH of Precipitation, July 1953—June 1954.

Presentation of the Observations

Samples of rain and/or snow collected by a small heated rain-gauge draining into a plastic sampling bottle are taken during each entire month at all the stations of the network. The data considered in this paper were collected during the period July 1953 through June 1954. During this period the net consisted of 19 stations, all within Sweden and with a maximum density in the southern half of the country. Starting in November 1954 the net became pan-Scandinavian in scope and was increased to a total of 44 stations. The data for the first two months of expanded operations will also be included in the presentation, as they emphasize certain points which are not so clearly illustrated by the smaller net.

The acidity or alkalinity of the samples has been determined by measurement of the electrical potential generated by the galvanic cell using a glass electrode and calomel electrode, with the sample as the electrolyte (Beckman pH meter, model G). The data are reported as pH units, *i.e.*, the negative of the common logarithm of the molar concentration of hydrogen ions in the sample.

It is customary in chemistry to take as neutral the pH value of 7.0 which is characteristic of ideally purified water at room temperature (25° C). However, the water in the atmosphere is in general not in this standard state of purity—in fact, it cannot possibly be. Even if all soluble solid electrolytes and pollution could be kept out of the cloud water, this water would always attempt to achieve equilibrium with the carbon dioxide pressure of the atmosphere. If one assumes that equilibrium is actually reached at the observed mean pressure of CO_2, the pH may be calculated from the known equilibrium constants to be 5.7 at 25° C. The variation with temperature is in the direct sense, although not linear.

Since the samples are measured at room temperature, this value of 5.7 may be regarded as the neutral point for atmospheric water. This convention will be adopted in this paper. Thus pH values lower than 5.7 represent a net acidity for the sample, while values greater than 5.7 indicate a net alkalinity.

The monthly values for all stations have been entered on synoptic charts and isopleths of pH have been drawn for each month. In addition, two kinds of "mean" values have been computed. In calculating these, the pH numbers are converted to their negative antilogarithms (molar concentrations of hydrogen ion); these are then multiplied by the measured amount of the sample. The sum is formed and divided by the sum of the sample volumes; this concentration is then reconverted to a pH value. It is apparent that this procedure, analogous to the method of calculating the mean molecular weight of a mixture, gives the only physically significant mean; a simple arithmetic mean of the pH values is quite without significance because of the variation in precipitation in space and time and the inherently nonlinear nature of the pH unit.

The "mean" obtained by the weighting system described above represents the pH which would be measured by an instrument if all the samples being averaged were poured together into a common vessel and no CO_2 were allowed to escape from the solution; put in another way, it represents the strength of a solution of a completely ionized acid which would have brought the same amount of hydrogen ion per unit area to the ground as was brought by the rainfall. The fact that more alkaline samples would partly neutralize more acid ones if physically mixed is disregarded, since a measure of total acid rather than net acid is the quantity desired. It is clear that the monthly samples actually underestimate the total acid in cases where individual rains fluctuate in pH because of the loss of CO_2 when fractions of different pH combine; thus the pH figures for each month are really slightly too high if interpreted as a measure of total acid brought to the ground.

The mean, or effective, pH for the entire net (areal mean) for each month, and the mean for the year at each station have been calculated by the procedure described above.

Fig. 2. pH of Precipitation, Sept. 1953.

Economy of page space does not permit the inclusion of the 15 charts which have been prepared. Figure 1 shows the distribution of the annual mean for each station for the year July 1953 to June 1954; this single chart displays most of the significant details which appear on the monthly charts. The two most atypical monthly charts are shown in Figs 2 and 3.

The principal features of the chart of Fig. 1 are: (a) the region of acid precipitation along the west coast; (b) the nearly neutral areas at the southwest tip of Sweden (Skåne) and along the southeast coast; (c) the acid tongue which separates these two neutral regions; (d) the acid region along the northeast (Bothnian) shore; (e) the large neutral or alkaline region covering the more elevated interior of Northern Sweden; (f) the acid "corridor" which joins the west-coast and Bothnian

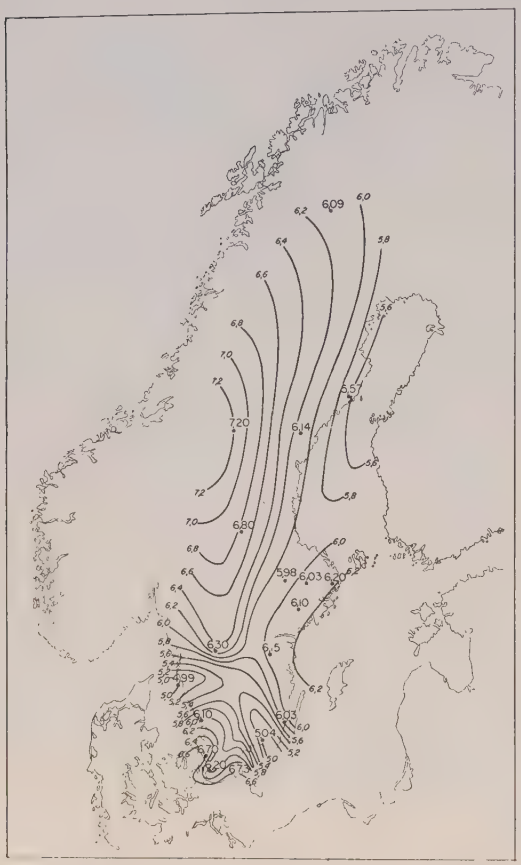


Fig. 3. pH of Precipitation, May 1954.

acid regions. This corridor shows a marked tendency to meander from month to month; in Fig. 2 it is shown displaced toward the northwest. When such a displacement occurs, a new region of alkalinity appears over the easternmost coastal region (Uppland). On other occasions the inland alkaline region extends southward to Uppland; the acid corridor then enters the Baltic farther south and the alkaline or neutral region of the southeast weakens strongly.

An examination of the monthly charts shows that the features discussed above are, in fact, semi-permanent features. The central values of pH fluctuate in magnitude from month to month and the positions of the centers shift slightly, but on the whole the pattern is remarkably persistent. The alkaline region of the northern interior (Norrland) does, however, show a distinct annual cycle; the pH is minimum there in late summer and is maximum in spring (compare Figs 2 and 3).

The monthly mean values for the entire network are shown plotted on the graph of Fig. 4. The solid curve applies to the values from the 19-station net for 1953—54, while the short dotted curve represents the means for the first two months of operation of the expanded 44 station net (November and December 1954). A distinct annual cycle is indicated, with a maximum of acidity (pH minimum) in the late fall and winter and a minimum acidity in the late spring. It is to be noted that, except for the month of May, the precipitation over Sweden as a whole is definitely more acid than the CO_2 equilibrium alone would permit; the neutral point is just reached in May. It is unfortunately true that the sampling in the alkaline north is sparse compared with the generally more acid south, so that allowance must be made for this in the interpretation; however, the curve would not be affected too strongly by the addition of more stations in the north because of the low winter precipitation there. The summer values would, however, be shifted upwards by the inclusion of more Norrland data. The two values for the new net (which has better northern coverage) show that the shift in the autumn means produced by doubling the number of stations is smaller than the month-to-month fluctuation.

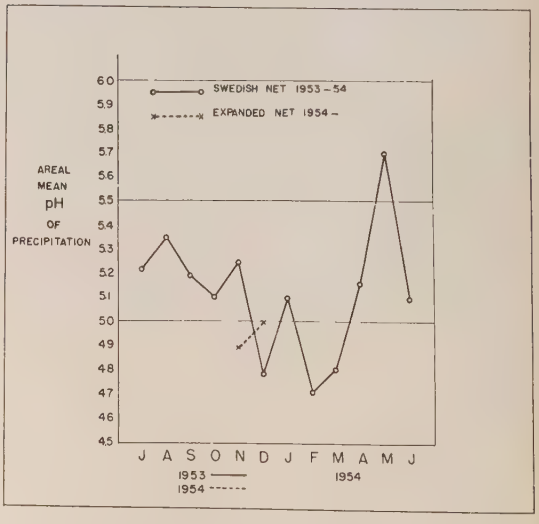


Fig. 4. Areal Mean of Precipitation pH vs. Time July 1953—June 1954 and Nov.—Dec. 1954.

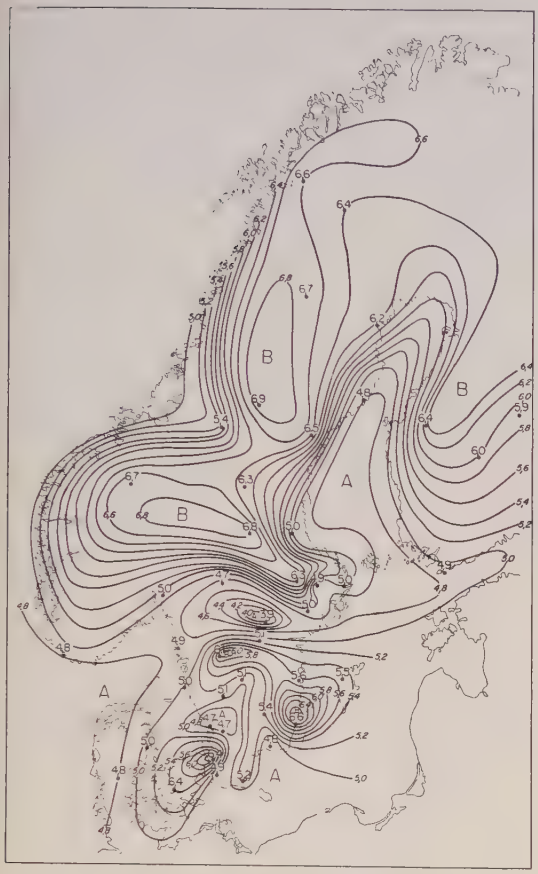


Fig. 5. pH of Precipitation—Expanded Network, November 1954.

Figs 5 and 6 show the pH distribution for the expanded network for the months of November and December 1954. It can be seen that the distribution is very much the same as for the smaller network; the acid character of most of the maritime regions and the alkaline character of the inland area are clearly illustrated. The acidity at the station in central Finland compared to the west Finnish station in the same latitude is, however, an anomalous feature.

Discussion

The data have shown that, on the whole, the precipitation over Sweden tends to be more acid than pure water in equilibrium with the carbon dioxide of the atmosphere. Furthermore, this excess acidity tends to be maximum at coastal stations and in the southern peninsula of Sweden, while the inland and northern regions tend to be neutral or slightly basic with respect to CO_2 equilibrium. The data from the larger network justify the extension of this picture to all of Scandinavia. The question of the source of the acid arises immediately. Examination of the anion analyses is not exceedingly helpful, since one cannot decide how much sulfate, for example, was collected as sulfuric acid and how much was picked up as neutral salts. One source which comes immediately to mind is combustion acids, primarily sulfuric acid, produced in densely populated and industrial regions. There is no particular difficulty in ascribing the acid tongue which enters Sweden from the south to this source, since the prevailing flow of air into this region is from the industrialized area of Central Europe during most rainy periods. The acid corridor across the southern peninsula of Sweden may also be attributed, at least in part, to such industrial acid. The acid region of the Sea and Bay of Bothnia may probably in part be due to the cellulose industries locates along the coasts in Norrland and Finland. One of the stations in the new network (Kvarntorp 1) is located in close proximity to a shale-oil plant; the pH values for this station are among the lowest recorded.

It appears difficult, on the other hand, to explain the high acidity of the precipitation on the south coast of Norway, the west coast of Sweden, the south coast of Finland and to some extent the coasts of the Sea and Bay of Bothnia as being caused by combustion products. It appears, therefore, that one must look to the sea itself as a source of acids. Provided that such a source can be found in the sea, the decreasing acidity with distance from the coast which is observed in areas free of pollution is readily explainable as a result of washing out of the acids by the precipitation, possibly coupled with addition of alkaline aerosols from the soil in snow-free months.

The next question is whether the sea can provide a source of volatile or easily-dispersed acid, and the chemical nature of this acid. Three possibilities immediately come to mind. The first of these is the possibility that organic acids which are products of biological activity may be concentrated in the thin surface film of the sea and dispersed from thence into the air by evaporation and/or the mechanical action of foaming and bubble breaking. This

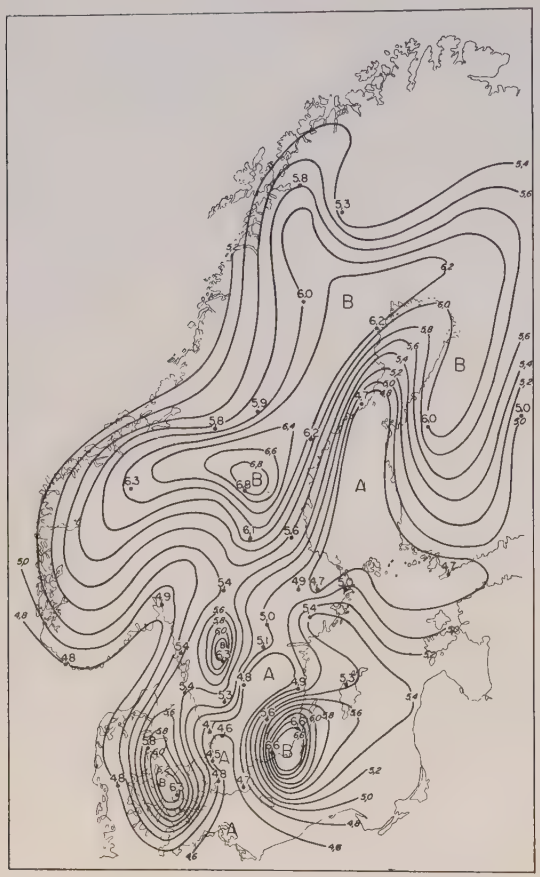


Fig. 6. pH of Precipitation—Expanded Network, December 1954.

possibility is amenable to experimental testing. Exactly the same argument applies to the hydrogen sulfide formed by biological action. This gas should be rather rapidly oxidized to SO_3 photocatalytically and by ozone, and would provide a supply of sulfuric acid. Unfortunately no data are apparently available on the rate of emission of H_2S from the sea, although it is known to occur. Presumably the gas is oxidized rather quickly, at least in daylight, and so the analyses of the air for sulfide would greatly underestimate the actual production. It remains to be determined whether or not this source strength is great enough to account for the observed acidity of the rainfall in coastal regions.

A third possible maritime source of acid may be sought in the ionic separation which has been observed to occur as a concomitant of the foaming of the sea surface (Cauer 1949, Köhler and Båth 1953). It is conceivable that the fine spray which ultimately becomes the source of many of the condensation nuclei is enriched in anions at the time of formation. This hypothesis is difficult to test, especially since the trapping and analysis of the smaller particles for anion-cation balance is no easy matter.

Although a decrease in acidity inland is to be expected as a result of the progressive washing-out of the (presumably maritime) acid contaminants as a precipitating air mass moves inland, it does not seem possible to account for pH values greater than about 5.7 in this way. In the mean, to be sure, the highest pH reported is 5.77 at Ås in central inland Sweden; however, the monthly values at the northern and inland stations are frequently above 6.0 and in a few cases above 7.0. It thus seems necessary to postulate the presence of alkaline aerosols of terrene origin. Examination of the concentrations of the various ions shows that many of these high pH values are associated with high calcium contents or a high ratio of calcium to sulfate; this suggests that aerosols of limestone are responsible for the alkalinity in the northern inland regions.

The two alkaline regions of Southern Sweden apparently result from a combination of pollution effects. The southeastern region coincides closely with the maximum area of surface limestone in Sweden; the Baltic islands of Gotland and Öland and the nearby Swedish coast possess considerable deposits of this mineral. The southwestern region, which also extends over eastern Denmark, is, in contrast to the area just mentioned, characterized by a high percentage of ammonium ions in most cases. The source of this ammonia is probably to be found in the extensive dairy farms of Denmark as a product of biological activity. However, a calcium pollution source exists at the southwest tip of Sweden in the form of a large cement factory; this point source frequently results in an anomalously high pH and calcium content at one or both of the two observing stations in this region. One may conclude, therefore, that the southwestern alkaline cell is clearly a result of human activity, in contrast to the southeastern cell and the northern inland.

The annual cycle of the network mean pH,

Fig. 4, is interpretable as a combination of effects. The minimum in February-March coincides rather well with the time of maximum combustion of fuel. At the same time the prevailing air flow during rain periods is from the south. In May the situation is reversed; the combustion decreases and the prevailing air flow is from the north and northeast in connection with the blocking anticyclone over the British Isles and the North Sea. In summer the blocking breaks down and the rain-bearing air may arrive from any direction; the maritime and industrial acids which are brought in from southerly and westerly areas bring the pH down from the spring maximum of 5.7 to intermediate values of around 5.2.

It is admitted that the making of such generalizations on the basis of only one year's data is slightly presumptuous. The conclusions are therefore put forward as tentative suggestions, subject to later revision when several years of data have been accumulated.

It would be highly desirable to attempt to relate the pH of precipitation to the region of origin of the associated air mass and to the past history of the air mass. It is unfortunately impossible to accomplish this in an unambiguous way if one is constrained to use total monthly samples of precipitation. One may attempt to relate the pH distribution to the mean isobaric pattern for the month and so obtain characteristic charts of pH distributions associated with air flows from various directions. These may, however, be somewhat misleading, since the precipitation samples represent, for particular stations, only a fraction of the days in the month. Furthermore, it sometimes happens that all of the rainfall at the station occurs on one or two days with completely atypical flow patterns. The monthly mean pressure pattern can therefore only be used for correlation studies when an examination of the daily charts shows that most of the precipitation did indeed arrive with wind patterns closely resembling the mean flow. A more satisfactory approach (although much more laborious) is to calculate precipitation-wind roses for each station and, by a study of the daily synoptic charts, to estimate trajectories of the air for each rain day and each station. In this way the fraction of the total precipitation which derives from a particular source region may be computed for each station.

A much better approach to the problem of correlation of precipitation properties to air masses would be to equip some stations (in regions free of local pollution) with large-aperture precipitation collectors and to analyse the samples collected at such stations during a single rain period. This procedure is contemplated, and it is expected that more reliable correlations between precipitation chemical composition and air masses will be obtained thereby. In the meantime the monthly mean isobaric patterns provide the only readily calculable information for correlation studies, but they must be used with caution.

REFERENCES

Cauer, H., 1949: Ergebnisse Chemisch-meteorologischer Forschung. *Arch. f. Meteor., Geophys. u. Bioklimat.*, Ser. B, **1**, 221—256.

Egnér, H., and Eriksson, E., 1955: Current data on the chemical composition of air and precipitation. *Tellus* **7**, 134—139.

Köhler, H., and Bâth, Markus, 1953: Quantitative chemical analysis of condensation nuclei from sea water. *Nova Acta Reg. Soc. Sci. Ups. Ser. IV, vol. 15, no.* **7**.

Rossby, C.-G., and Egnér, H., 1955: On the chemical climate and its variation with the atmospheric circulation pattern. *Tellus* **7**, 118—133.

Microdetermination of CO_2 in the Air, with Current Data for Scandinavia

STIG FONSELIUS, Institute of Meteorology, University of Stockholm, Sweden
FOLKE KOROLEFF, Institute of Marine Research, Helsingfors, Finland
with Preface by Professor KURT BUCH, Finland

(Manuscript received March 16, 1955)

Abstract

Krogh and Brandt Rehbergs method for estimation of CO_2 in atmospheric air has been investigated and modified. The method is used for continous investigation of the CO_2 content of the air in Scandinavia. A permanent net of sampling stations has been established, 6 in Sweden, 4 in Finland, 3 in Norway and 2 in Denmark. The samples are taken three times each month. A simple method of taking air samples is described. Current data for November, December 1954 and January 1955 are published.

Preface

The problem of the carbon dioxide content of the free atmosphere and its variations has recently acquired a renewed importance. Callendar (1940) published a critical survey of the entire literature relating to CO_2 determinations for the period from 1860 to the middle of the 1930's, in which he asserted that the most recent analyses show higher concentrations than the older determinations made before and at the turn of the century. The increase amounts to about 10 per cent on the average. Since, in this survey of the material, only values which were reliable from the standpoint of precision in analytical procedure were considered, and since the mean values arrived at by various investigators are in good agreement with each other, Callendar regards the increase as real. He ascribes the increase to the enormously expanded industrial consumption of fuel since the beginning of the century. He finds support for this conclusion in the observation that the total consumption of industrial fuel from 1900 to 1935 (calculated from industrial statistics) has supplied to the atmosphere a total amount of CO_2 which is in satisfactory agreement with the quantity which can be calculated from the aforementioned 10 per cent increment arrived at on the basis of the CO_2 analyses.

The principal objection which can be raised against this conclusion is that the atmosphere is involved in steady gas exchange with the sea. If one assumes that the sea and the atmosphere have achieved equilibrium with each other as a result of exchange processes operating over thousands of years, then any sudden increase in atmospheric CO_2 should be smoothed out through absorption of the excess by the sea. Callendar is also aware of this possibility, but considers that this process goes on with extreme slowness, perhaps during several thousand years, because of the sluggish vertical circulation in the sea, of whose total volume only a small fraction (about 1.3 per cent) can possibly come into immediate contact with the atmosphere at any given time.

The question of this exchange process may be regarded as a separate problem with its own intrinsic interest, but it certainly stands in an intimate connection with the question of a possible increase in the CO_2 content of the atmosphere. A settlement of this question requires a large volume of observational data distributed over an extensive region of space and time. As twenty years have now passed since the last series of observations by J. Haldane (1936) and K. Buch (1934, 1939 a, 1939 b) (on which Callendar in large measure based his conclusions) were published, and because the CO_2 content of the atmosphere may have been increased further since that time, it seems especially desirable to begin the collection of new series of observations, preferably over as large an area of the earth's surface as possible.

A suitable opportunity to realize this wish presented itself in connection with the plans for a world-wide investigation of "trace elements" in the atmosphere, which were drawn up on the initiative of Professor C.-G. Rossby, director of the Institute of Meteorology, Stockholm University. A resolution to begin these investigations was approved at a conference in Stockholm in May 1954 (Eriksson, 1954), which was attended by representatives from various parts of the world. With reference to CO_2, it was resolved to set up a regular sampling program at several stations in all these countries participating in the effort. I undertook the responsibility for organizing Finland's share of the project.

The locations of the sampling stations, the method of sampling and the analytical procedures are described in subsequent sections of this paper.

Sampling technique

The air samples are taken on open places without foliage, such as a lake- or sea-shore, or on a rock without forest. Towers, or roofs without chimneys, are also suitable places. The samples are always taken at the same place three times each month, on the 1st, the 10th and the 20th, at about 13.00 hours.

The samples are taken in gas pipettes, which have two stopcocks and a capacity of about 250 cc. A rubber bulb is connected to the pipette (figure 1) and both the stopcocks are opened. The vent of the rubber bulb is turned towards the wind direction as far as possible from the sampler (figure 2). 150—200 pumpings with the bulb are rapidly carried out, to get a complete air exchange. Then stopcock (a) is closed and two pumpings on the bulb are carried out to get an overpressure in the pipette. Then stopcock (b) is closed and the bulb removed. The samples are sent to the laboratories for analysis, with a report of the weather conditions.

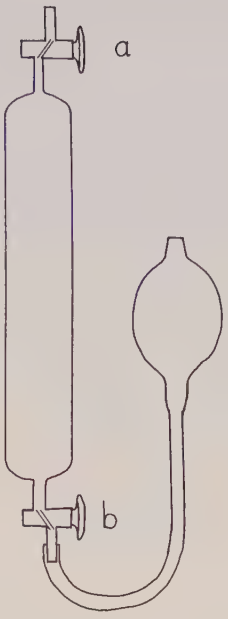


Fig. 1. Sampling Pipette with Rubber Bulb.

Fig. 2. Mr. Roland, Plönninge taking an air sample.

The samples are being taken at the following places (see fig. 3):

Sweden

Abisko Research Station of Academy of Science, by the lake Torne Träsk (on the roof of the building)
Öjebyn Experiment Station, near Piteå (on a gravelled plain)
Bredkälen, village, near Strömsund (on the roof of a barn)
Ultuna Agriculture College, near Uppsala (on a gravelled plain)
Flahult, Experiment Station, near Jönköping (on a gravelled plain)
Plönninge Agricultural school, near Halmstad (on a gravelled plain)

Finland

Kauhava Military Training school (on the airfield)
Rissala aerodrome, near Kuopio (on the airfield)
Luonetjärvi aerodrome, near Jyväskylä (the flight control tower)
Tvärminne Zoological Station, near Hangö (on the seashore)

Norway

Vågönäs Experiment Station, near Bodö (on the railway embankment)
Vågåmo, Klones Agricultural School (at the shore of the lake Vågåvatn)
Bergen, University of Bergen (the tower of the university)

Denmark

Ödum Agricultural Experiment Station, near Vejen (on gravelled ground)
Askov Agricultural Experiment Station, near Århus (on gravelled ground)

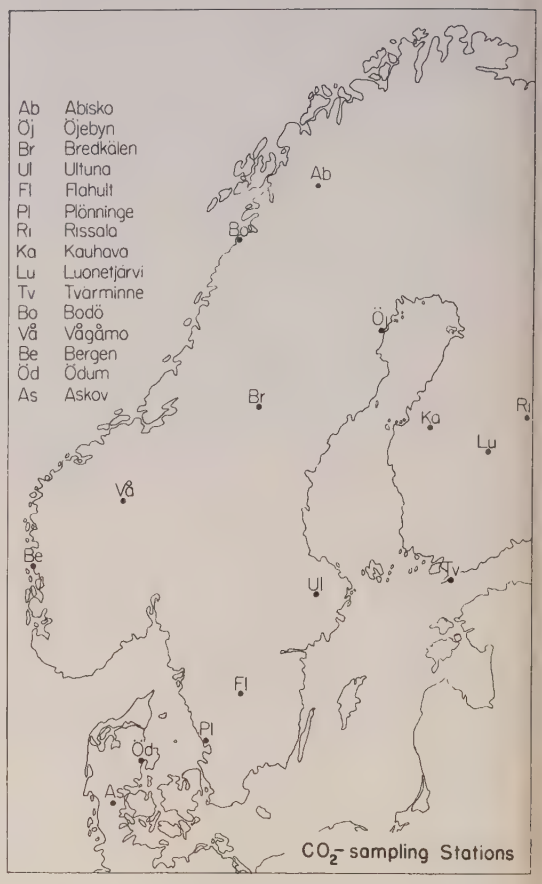


Fig. 3. Map showing the distribution of stations.

Method of analysis

The method developed by KROGH and BRANDT REHBERG (1929) was used as a basis for the determination of the amount of carbon dioxide in atmospheric air. In brief, the principle of this method is to pass a definite small amount of air through a solution of barium hydroxide and then to titrate the excess of the hydroxide with hydrochloric acid.

Absorption tube. The tube has the form shown in figure 4. It is made of Pyrex glass, has an inner diameter of about 4 mm, and its lower end is drawn into a slowly tapering point (1 mm) bent upwards semicircularly. At a distance of 1.5 cm from the upper end, the tube has been blown to a bulb with 1.5 cm diameter. The total length of the tube is about 15 cm.

Apparatus. The levelling bulb A contains mercury, and G and H are three-way stopcocks. The air sample pipette B is connected between G and H. The second branch of the stopcock H is connected to an adjustable CO_2-free air stream, obtained by passing air through a sufficiently long tube E filled with Ascarite or Soda lime. The third branch is drawn out

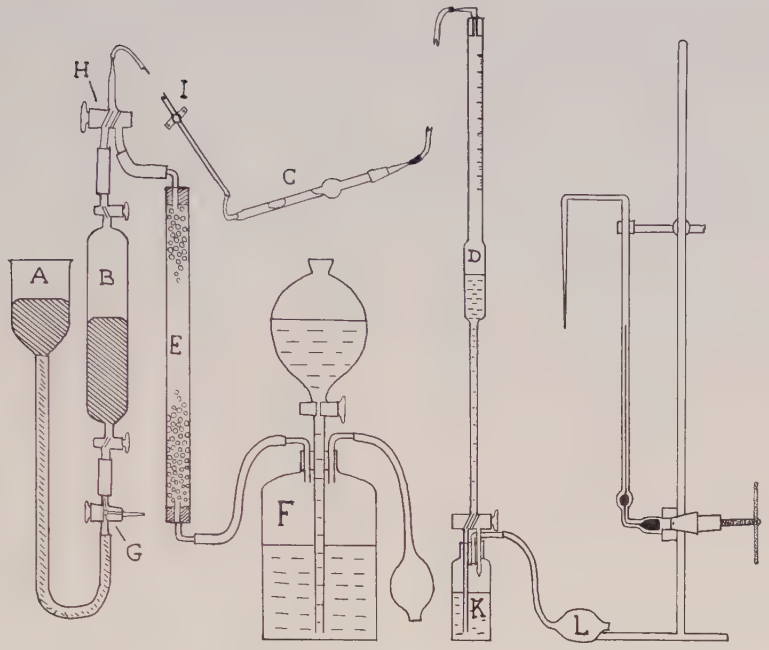


Fig. 4. CO_2-Absorption Apparatus with Microtitration Pipette.

into a capillary and connected to the capillary end of the absorption tube C with a thin pressure tube of inner diameter 1.5 mm and outer 3 mm. A pinch clamp J is placed on the tubing to regulate the rate of flow. The other end of C is connected to the measuring burette D with a similar pressure tube. The burette can have a capacity of 60 or 100 ml and is filled with water by compressing air in bottle K. (Fig. 4.)

Hydrochloric acid, 0.08956 N solution. One ml of this acid is equivalent to one ml of CO_2 at NTP. (The mol. weight of CO_2 = 44.01 and its liter weight = 1.9769 grams.)

Barium hydroxide, 0.0179 or 0.0119 N solution. If the capacity of the microburette for HCl is 0.2 ml, the stronger hydroxide is used. The other normality is used if the capacity of the burette is 0.1 ml. The solution should contain 0.5 per cent of $BaCl_2$ and be stored in bottles protected by an Ascarite tube. (Fig. 5.)

Titration of the hydroxide with the acid. The microburette for the acid, capacity 0.1 ml, is provided at the top with a U-shaped glass tip drawn into a capillary. The acid is drawn in through this capillary using a mercury and screw piston in the other end of the burette. Carbon dioxide-free air is blown through the absorption tube held in a tilted position for a few minutes, the flow being finally reduced to a very slow rate. The tube is now filled with 0.75 ml of the weaker hydroxide using an all-glass tuberculine syringe containing one ml with a scale divided into 100 units and equipped with a long glass tip ending

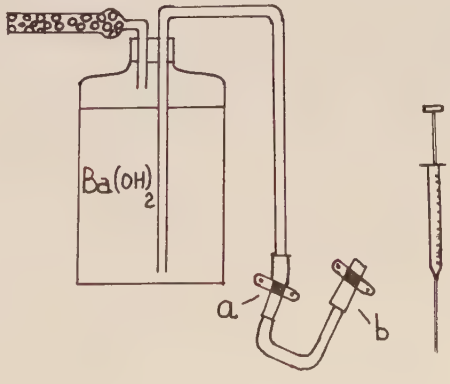


Fig. 5. Store Bottle for $Ba(OH)_2$ solution.

in a capillary. The storage bottle for the hydroxide (fig. 5) is equipped with a syphon tube and U-shaped termination. A pinch clamp is located between them. The U-tube is equipped with a rubber tube and a second pinch clamp. This second clamp is first removed and the first clamp is opened so as to allow fresh hydroxide to flow and flush out the older solution from the U-tube. The first clamp is then closed; the syringe is inserted into the open end of the U-tube, washed out once with hydroxide and then filled in the usual manner. With this technique, the aliquot taken is protected from atmospheric carbon dioxide.

When the absorption tube is filled, one drop of a 0.1 % cresol- or phenolphthalein solution is added and the titration carried out with the tip of the burette inserted into the tube below the bulb. During the titration a slow CO_2-free air stream flows through the hydroxide. Two blank titrations are not allowed to differ more than half a scale unit or 0.0005 ml from each other.

Procedure

The analysis is begun by attaching the pipette between G and H. (See fig. 4.) The cock G is first turned so as to connect the bulb A to the atmosphere. Now H is turned so that the through connection is from the pipette to the pressure tube. (The cocks of the pipette are closed.) Bulb A is raised until mercury begins to flow out of G, and this cock is then turned so as to connect the pipette to the atmosphere; the lower stopcock of the pipette is then opened so that the slight over-pressure in the pipette flushes the air out of the short length of connection tubing. At this stage G is turned so as to connect the pipette to A. The bulb is raised so as to provide an over-pressure in the pipette. The upper cock of the pipette is opened and the air is flushed out of the cock H and the rubber tubing. (At this stage the absorption tube is not yet filled with hydroxide and is not connected to the measuring burette D.) Thereafter, H is turned so that CO_2-free air can flow through the absorption tube for a few minutes. Burette D is filled to the mark with water, and hydroxide filled into C in the manner already described under titration. Next, the absorption tube is connected to burette D and cock H is turned so as to connect the pipette to the tubing leading to C. The cock of burette D is then turned so that D may drain back into Reservoir K. The rate of flow is adjusted with clamp J and also by turning the tube to a suitable tilted position. The flow is correct when the small air bubbles follow each other as pearls on a string. With this arrangement every air bubble is in contact with the hydroxide 20 to 30 seconds, and the time required to drain D (100 ml) is approximately one hour. When the main bulb of the burette is empty, the levels of mercury in A and B should be adjusted to be nearly at the same height to avoid over-pressure at the end of the absorption.

After burette D has drained completely, cock H is again connected to the CO_2-free air flow and the air is allowed to flow slowly during the subsequent titration. The absorption tube is disconnected from the measuring burette and the hydroxide titrated, as already described under titration.

To avoid appreciable changes in temperature, direct sunlight has to be avoided. During the absorption, pressure and temperature are read and the volume of the measuring burette reduced to NTP.

After an analysis the absorption tube can be used for another absorption without washing. If the tube must be washed, a blank titration is necessary.

Calculations:

1 mm^3 HCl is equal to:

$$\frac{\text{normality of HCl} \cdot 44.01/2}{1.9769} \ mm^3 \ CO_2 = \alpha$$

Blank titration a mm^3 of HCl

After an absorption b mm^3 of HCl

Pressure during the absorption $= p_1$

Temperature » » » $= T_1$

Volume of measuring burette $= v_1 mm_3$

Volume of the air sample at

$$\text{NTP} = v_0 = \frac{v_1 \cdot T_0}{P_0} \cdot \frac{p_1}{T_1}$$

$$CO_2\,\% = \frac{100\,(a-b)\,x \cdot \alpha}{v_0}$$

As mentioned above, the time of absorption for 100 ml is stated to be about one hour, and each small air bubble must be in contact with the hydroxide for 20 to 30 seconds. Krogh and Brandt Rehberg stated that 20 minutes

was sufficient for absorbing the carbon dioxide from 100 ml. An investigation of the effect of the absorption time gave the results shown in Table 1.

Table 1. Effect of time on the absorption 100 cm^3 air sample, blank titration 98.0 mm^3

Time in min.	HCl mm^3 aft. absorp.	CO_2 cm^3/10 l
19	71.0	2.89
24	68.0	3.18
39	66.0	3.41
47	65.5	3.47
71	65.1	3.51
85	65.1	3.51

As can be seen from the table, an absorption time of about an hour is necessary. It can be mentioned also that the values given by Buch (l. c.) are based on an absorption time of this length; therefore his values can be directly compared with our own.

As the analyses have been carried out partly in Stockholm (Mr. Wärme) and partly in Helsingfors (Koroleff), systematic errors may exist in the analysis methods. To test this possibility, double samples have been taken; the first analyzed in Stockholm and the second in Helsingfors, and vice versa. However, the values in Table II show that a fully satisfactory agreement exists.

Table II

	Stockholm	Helsingfors
	CO_2CC/10 l	
1	3.30	3.25
2	3.48	3.51
3	3.40	3.47

In Table III, the CO_2 values for Nov. 1954—Jan. 1955 are given in cc per 10 l of air. As the variations of the CO_2 content in atmospheric air might be of interest to geophysicists and meteorologists, the CO_2 values will be published quarterly in this journal.

These investigations have been financed mainly by the Wallenberg Foundation in Sweden and the State Scientific Commission in Finland.

Table III. CO_2 values November 1954—January 1955

November	Date	Hour	t °C	Wind m/sec.	Weather and Pressure	CO_2 cc / 10 l
Ab.	1	13	— 2	Calm		3.66
	10	—	—	—		—
	20	13	—11	SSE 5.1		3.59
Öj.	1	13	—	NE 3		(4.22)
	10	13	— 1	NE 10	Snow	3.11
	20	13	— 6	NW 5	»	3.20
Br.	1	13	+ 4	NW 3		(4.29)
	10	—	—	—		—
	20	13	— 9	N 3		3.26
Ul.	1	13	+ 8	N 3		3.23
	10	13	+ 2	N 3		3.61
	20	13	— 1	N 3		3.17
Fl.	1	—	—	—		—
	10	—	—	—		—
	20	09	— 2	N 3		3.52
Pl.	1	14	+10	SW 3		(3.95)
	10	13	+ 7	W 10		3.31
	20	—	—	—		—
Ka.	1	—	—	—		—
	10	13	0	SSW	744.9	3.41
	20	13	— 5	E	775.6	2.73
Ri.	1	—	—	—	—	—
	10	16	+ 3	SE	749.9	2.92
	20	16	— 5	NE	770.5	3.09
Lu.	1	—	—	—		—
	10	12	+ 1	SSE	735.1	3.89
	20	12	— 5	ENE	764.5	3.40
Tv.	1	—	—	—	—	—
	10	09	—	—	785.0	3.12
	20	12	— 1	NE	778.0	3.14
Bo.	1	13	+ 3	N 5		3.75
	10	13	+ 1	S 5		3.05
	20	13	+ 4	SSE 6		3.62
Vå.	1	13	—	SW 3		3.13
	10	13	— 6	NW 3		3.85
	20	13	—	NW 2		2.28
Be.	1	—	—	— 3		—
	10	13	+ 6	SSE 3	Rain	3.44
	20	—	—	— 3	»	—
Öj.	1	16	—	—		3.36
	10	14	+ 7	W 10		3.49
	20	13	+ 2	SE 15		3.39
As.	1	14	+ 9	SSW 3		3.36
	10	13	+ 8	SW 10		3.15
	20	13	+ 2	SE 1		3.46

Table III (Cont.)

	December						January					
	Date	Hour	t °C	Wind m/sec.	Weather and Pressure	CO_2 cc / 10 l	Date	Hour	t °C	Wind m/sec.	Weather and Pressure	CO_2 cc / 10 l
Ab.	30	13	0	ENE 13		3.24	30	13	—12	SSW 4		3.29
	10	13	— 5	ESE 18	Snow	3.19	10	13	—20	Calm		3.46
	20	13	— 1	WSW 5	»	3.36	20	18	—18	S 3		3.58
Öj.	1	13	— 1	SE 3		3.16	1	14	+ 1	NW 10		3.29
	10	13	0	E 10	Snow	3.32	10	13	— 2	S 3		3.55
	20	13	— 3	Calm		3.23	20	13	— 6	NE 3		3.50
Br.	1	13	— 1	SSE 15		3.34	1	13	+ 1	W 3		3.40
	10	13	— 1	S 15	Snow	3.74	10	13	—15	NE 5	Snow	3.49
	20	—	—	—	—	—	20	13	— 8	W 3		3.63
Ul.	1	13	+ 2	SSE 3	Rain	3.60	1	13	— 4	SSW 1		3.24
	10	—	—	—	—	—	10	14	— 3	SSW —		3.82
	20	14	+ 5	SW 3		3.68	20	13	— 9	Calm		3.63
Fl.	1	13	+ 3	S 3	Rain	3.39	1	13	— 3	Calm		3.17
	10	13	+ 4	S 3		3.83	10	13	— 4	SW 10	Snow	3.55
	20	13	+ 5	SW 10	Rain	3.22	20	13	— 5	NE 3		3.36
Pl.	1	13	+ 6	Calm		3.39	1	13	— 1	NE 5		3.26
	10	13	+ 5	S 10		3.33	10	13	— 3	S 5	Snow	3.27
	20	13	+ 5	W 5		3.19	20	13	— 5	Calm		3.42
Ka.	1	13	— 2	SE 9	764.9	3.18	1	13	— 5	SSW 5	774.5	2.94
	10	13	— 1	SE 16	751.8	2.87	10	13	— 1	S 5	742.5	3.28
	20	13	0	S 9	733.9	3.24	20	—	—	—	—	—
Ri.	1	14	— 6	SE 4	764.9	3.11	1	14	—15	SE 3	771.9	3.24
	10	—	—	—	—	—	10	15	— 1	S 5	740.5	2.98
	20	15	0	SSE 5	728.1	3.79	20	14	—15	WNW 2	739.3	2.95
Lu.	1	13	— 4	SE 6	756.5	3.06	1	—	—	—	—	—
	10	13	— 3	ESE 8	744.9	3.36	10	12	— 1	S 5	735.6	3.65
	20	12	0	SSE 4	726.5	3.25	20	12	— 9	NW 3	731.6	3.57
Tv.	1	12	— 1	SE —	778.0	3.16	1	12	— 3	NE —	779.0	3.02
	10	12	+ 2	S —	768.0	3.17	10	12	0	SW 12	765.0	3.17
	20	—	—	—	—	—	20	13	— 1	W 2	760.0	2.97
Bo.	1	13	+ 5	SSW 4		3.45	1	13	+ 4	WNW 3		3.43
	10	13	+ 2	ENE 5		3.53	10	13	— 7	NE 3		3.52
	20	13	— 1	NNE 2		3.62	20	13	— 7	NE 3		3.46
Vå.	1	13	+ 2	SE 3		3.27	1	13	— 5	Calm		3.46
	10	13	— 1	ENE 2	Snow	3.17	10	13	— 2	NE 10	Snow	3.52
	20	13	— 1	NE 5		3.22	20	13	— 3	W 10		3.50
Be.	1	13	+10	SSE 3		3.49	1	13	+ 4	Calm		3.36
	10	13	+ 6	ESE 1		3.09	10	13	+ 2	SSE 1	Snow	3.70
	20	13	+ 3	NNW 1	Snow	3.61	20	13	— 1	ESE 3		3.53
Öj.	1	13	+ 6	S 3		3.34	1	13	— 1	NE 3		3.13
	10	—	—	— 10		—	10	13	— 1	SSW 10	Snow	3.69
	20	13	+ 7	SW 10		3.23	20	13	— 4	Calm		3.53
As.	1	13	+ 7	S 3		3.24	1	13	0	NE 3		3.25
	10	—	—	—		—	1	13	0	NE 3		3.57
	20	—	—	—		—	20	13	— 2	NW 3		3.76

REFERENCES

Buch, K., 1934: Beobachtungen über chemische Faktoren in der Nordsee, zwischen Nordsee und Island, sowie auf dem Schelfgebiete nördlich von Island. *Conseil perm. intern. p. l'exploration de la mer. Rapp. et proc. verb.*, **89**, 13.

— 1939: Kohlensäure in Atmosphäre und Meer an der Grenze zum Arktikum. *Acta Acad. Aboensis, Mat. et Phys.* XI, 12.

— 1939: Beobachtungen über das Kohlensäuregleichgewicht und über den Kohlensäureaustausch zwischen Atmosphäre und Meer im Nordatlantischen Ozean. *Ibid.*, XI, 9.

Callendar, G. S., 1940: Variations of the Amount of Carbon Dioxide in Different Air Currents. *Quarterly J. of the Meteorol. Soc.*, **66**, 278.

Eriksson, E., 1954: Report on an Informal Conference in Atmospheric Chemistry Held at the Meteorological Institute, University of Stockholm, May 24—26 1954. *Tellus*, **6**, 302.

Haldane, J. B. S., 1936: *Nature* **137**, 575, cit. Callendar 1940.

Krogh, A., und Brandt, Rehberg, P., 1929: CO_2-Bestimmungen in der Atmosphärischen Luft durch Mikrotitration. *Biochemische Zeitschr.* **205**, 265.

NOTES

Current Data on the Chemical Composition of Air and Precipitation

(For further information see Egnér, H., Eriksson E., *Tellus* **7**, 134—139, 1955.)

D 501

Station	Code	mm	mg/m²								pH	$\frac{\mu val}{l}$ HCO_3^-	ohm^{-1} cm^{-1} $\varkappa \cdot 10^6$
			S	Cl	NO_3-N	NH_3-N	Na	K	Mg	Ca			
Riksgränsen	Ri	11	12	77	1	1	58	2	7	17	6,6	34	45
Kiruna	Ki	23	6	16	1	0	11	0	2	3	5,8	31	6
Arjeplog	Ar	20	6	35	3	2	19	2	5	29	6,6	39	18
Öjebyn	Öj	32	22	18	5	8	16	7	5	21	6,0	9	12
Röbäcksdalen	Rö	19	20	22	3	2	29	2	5	22	6,6	40	20
Offer	Of	28	25	117	6	15	93	28	6	22	6,8	74	33
Bredkälen	Br	6	8	24	3	3	15	0	9	0	6,3	4	29
Åre	ÅF	52	19	46	3	5	30	2	9	6	5,4	0	7
Åre, Hummeln	ÅH	—	—	—	—	—	—	—	—	—	—	—	—
Sveg	Sv	22	20	7	5	4	8	2	3	13	5,3	0	13
Mora	Mo	12	29	9	5	11	5	5	10	32	6,7	45	41
Åmot	Åm	21	13	11	5	3	21	4	5	33	6,7	55	15
Sala	Sa	26	47	22	7	7	24	7	7	48	5,4	0	30
Ultuna	Ul	29	31	17	10	13	12	5	9	7	4,6	0	23
Erken	Er	25	27	23	7	7	16	10	4	16	5,0	0	19
Strängnäs	St	16	25	15	7	12	10	1	6	13	4,8	0	35
Forshult	Fo	33	27	14	7	8	20	2	8	14	5,3	0	13
Kvarntorp	Kv1	10	156	28	9	63	30	23	16	410	3,9	0	345
Kvarntorp	Kv7	10	37	36	7	16	13	2	12	28	4,7	0	58
Västra Ny	VN	18	26	19	6	5	23	3	5	34	6,2	34	21
Lanna	La	13	29	33	10	15	30	3	9	28	6,5	57	40
Bornö	Bo	34	49	199	18	13	143	9	22	50	6,2	38	42
Vinga	Vi	11	79	1 420	15	9	836	39	111	57	4,5	0	505
Falsterbobruk	Fa	46	50	40	10	12	32	4	7	14	4,6	0	23
Flahult	Fl	14	21	42	4	3	33	5	7	29	6,8	98	35
Ambjörnarp	Am	37	43	99	13	14	72	32	9	26	6,6	62	27
Simlångsdalen	Si	58	67	69	20	30	45	7	17	21	4,7	0	24
Plönninge	Pl	15	57	154	17	44	101	9	14	29	6,3	32	86
Söraby	Sö	14	27	31	5	5	40	4	4	29	6,7	80	35
Smedby	Sm	12	15	23	3	3	19	3	1	7	5,3	0	20
Sylfaste	Sy	21	36	34	9	9	25	3	2	30	4,9	0	26
Bräkne-Hoby	BH	13	22	39	8	14	30	2	4	25	5,9	9	40
Hammenhög	Ha	11	41	88	9	20	65	6	10	40	6,5	43	74
Alnarp	Al	24	100	192	19	25	121	9	17	115	5,3	0	77
Hilleshög	Hi	7	34	86	9	17	54	4	8	19	4,5	0	101
Ås	Ås	22	42	65	13	19	59	7	9	35	6,0	18	36
Vågåmo	Vå	9	7	7	2	3	9	2	0	15	6,5	24	16
Lista	Li	56	127	2 340	33	26	1 370	62	172	75	4,4	0	188
Kauhava	Ka	16	17	12	3	5	9	4	1	12	5,9	0	16
Kuopio	Ku	28	17	9	4	3	11	3	0	11	5,4	0	9
Jyväskylä	Jy	21	14	15	5	3	15	4	3	12	5,7	10	13
Tvärminne	Tv	26	35	51	10	9	37	6	6	12	4,4	0	33
Ödum	Öd	10	36	124	6	20	74	6	11	27	6,7	67	86
Askov	As	22	74	538	11	25	333	13	38	39	4,9	0	128
Tystofte	Ty	11	55	149	11	30	101	6	19	45	5,7	58	114

D 502

Station	Code	mm	mg/m²								pH	$\frac{\mu val}{l}$ HCO_3^-	ohm⁻¹ cm⁻¹ $\varkappa \cdot 10^6$
			S	Cl	NO_3-N	NH_3-N	Na	K	Mg	Ca			
Riksgränsen	iR	33	7	23	0	0	17	4	5	4	5,6	7	6
Kiruna	K	37	9	0	2	0	5	4	1	2	5,2	0	6
Arjeplog	Ar	24	8	13	3	0	12	3	3	17	6,2	10	11
Öjebyn	Öj	28	24	12	6	7	29	9	7	30	6,6	45	18
Röbäcksdalen	Rö	22	28	6	7	7	9	5	2	11	4,6	0	24
Offer	Of	9	15	7	4	4	13	12	1	12	6,0	34	26
Bredkälen	Br	27	21	21	5	12	19	21	1	5	5,8	2	14
Åre	ÅF	34	15	28	4	3	19	11	7	11	5,5	0	10
Åre, Hummeln	ÅH	5	10	3	2	0	8	4	1	3	5,7	30	28
Sveg	Sv	26	8	1	2	0	3	3	0	1	5,1	0	8
Mora	Mo	14	21	12	3	5	6	15	3	32	6,6	50	25
Åmot	Åm	27	18	10	4	3	22	8	3	19	6,5	28	12
Sala	Sa	24	33	13	5	6	31	11	4	43	6,8	58	24
Ultuna	Ul	12	22	6	7	11	8	4	3	5	4,4	0	37
Erken	Er	13	22	10	6	4	7	4	7	10	4,3	0	39
Strängnäs	St	18	15	11	7	7	11	10	7	20	5,3	0	20
Forshult	Fo	16	12	5	5	3	13	4	2	10	6,1	26	13
Kvarntorp	Kv1	6	104	9	4	10	13	9	4	127	4,5	0	192
Kvarntorp	Kv7	19	34	11	8	10	8	6	2	24	4,6	0	32
Västra Ny	VN	18	23	17	9	7	25	8	3	25	6,3	30	24
Lanna	La	7	13	10	2	4	13	2	2	13	6,8	55	30
Bornö	Bo	9	13	17	6	3	30	4	8	19	—	109	44
Vinga	Vi	9	21	113	7	7	72	7	13	13	5,3	0	85
Falsterbobruk	Fa	37	41	30	19	13	26	7	5	19	4,5	0	30
Flahult	Fl	22	22	16	7	4	11	7	4	18	5,1	0	18
Ambjörnarp	Am	23	17	28	8	4	23	8	3	26	6,2	21	18
Simlångsdalen	Si	30	22	35	11	6	32	9	4	21	5,3	0	17
Plönninge	Pl	12	40	52	12	28	51	5	6	29	6,4	40	66
Söraby	Sö	14	19	12	8	5	22	4	3	18	6,2	47	24
Smedby	Sm	26	42	19	17	18	15	6	3	12	4,5	0	36
Sylfaste	Sy	22	33	15	14	12	18	5	3	18	4,7	0	31
Bräkne-Hoby	BH	25	42	29	22	24	19	7	6	15	4,4	0	43
Hammenhög	Ha	12	26	20	9	14	17	5	3	12	4,8	0	41
Alnarp	Al	15	34	33	9	11	18	6	16	32	4,9	0	43
Hilleshög	Hi	5	35	21	6	7	42	15	4	27	7,0	204	124
Ås	Ås	30	34	15	12	11	33	7	5	28	6,2	15	20
Vågåmo	Vå	0,4	—	—	9	17	—	—	—	—	—	—	—
Lista	Li	48	88	704	27	18	429	29	52	43	4,4	0	96
Kauhava	Ka	15	11	4	3	5	8	4	2	9	6,2	29	15
Kuopio	Ku	9	7	3	2	1	6	4	3	7	6,2	13	15
Jyväskylä	Jy	14	11	3	3	2	7	4	3	7	5,4	0	13
Tvärminne	Tv	7	19	6	5	6	10	3	3	9	4,6	0	48
Ödum	Öd	13	29	29	8	16	22	5	8	10	4,5	0	51
Askov	As	12	25	47	7	19	32	7	2	13	4,8	0	52
Tystofte	Ty	11	50	37	8	35	41	5	4	24	6,6	59	77

D 503

Station	Code	mm	mg/m²								pH	$\frac{\mu val}{l}$ HCO_3^-	ohm^{-1} cm^{-1} $\varkappa \cdot 10^6$
			S	Cl	NO_3-N	NH_3-N	Na	K	Mg	Ca			
Riksgränsen	Ri	37	19	77	2	3	46	6	6	16	6,3	16	15
Kiruna	Ki	30	6	11	0	0	8	3	2	8	6,0	20	5
Arjeplog	Ar	12	7	10	1	1	9	2	2	9	6,6	74	14
Öjebyn	Öj	5	9	3	1	2	4	4	2	10	6,2	94	28
Röbäcksdalen	Rö	0	0	0	0	0	0	0	0	0	0	0	0
Offer	Of	6	15	17	1	3	23	5	3	17	7,0	149	46
Bredkälen	Br	0	0	0	0	0	0	0	0	0	0	0	0
Åre	ÅF	19	14	32	1	2	31	3	4	13	6,5	35	19
Åre, Hummeln	ÅH	0	0	0	0	0	0	0	0	0	0	0	0
Sveg	Sv	0,8	—	—	—	—	—	—	—	—	7,7	800	156
Mora	Mo	0,1	—	—	—	—	—	—	—	—	—	—	—
Åmot	Åm	1,5	—	—	—	—	—	—	—	—	7,4	308	98
Sala	Sa	3	23	10	2	4	27	8	5	34	7,4	740	128
Ultuna	Ul	4	12	4	2	4	3	2	1	6	5,2	0	41
Erken	Er	9	16	5	2	2	5	3	1	9	4,6	0	31
Strängnäs	St	4	20	9	3	2	19	6	4	25	7,1	275	83
Forshult	Fo	3	7	5	2	0	20	3	2	16	7,5	446	70
Kvarntorp	Kv1	12	174	16	5	10	26	16	10	289	6,5	76	162
Kvarntorp	Kv7	13	34	10	6	8	6	5	4	29	5,2	0	36
Västra Ny	VN	17	19	8	4	2	15	3	3	24	6,6	59	21
Lanna	La	5	17	17	2	3	16	4	3	21	7,0	145	60
Bornö	Bo	8	18	51	6	4	33	5	6	11	5,8	33	49
Vinga	Vi	10	58	558	12	6	334	19	43	30	4,8	0	282
Falsterbobruk	Fa	33	50	22	8	11	19	8	4	17	4,5	0	30
Flahult	Fl	24	25	13	5	3	16	4	4	20	6,0	19	15
Ambjörnarp	Am	19	23	24	5	3	24	7	4	27	6,5	47	24
Simlångsdalen	Si	67	49	67	13	11	40	6	6	12	4,7	0	21
Plönninge	Pl	36	82	71	10	62	51	8	6	20	5,5	0	36
Söraby	Sö	29	37	26	6	9	18	5	4	32	5,7	10	20
Smedby	Sm	29	58	34	10	17	60	10	7	44	6,5	40	33
Sylfaste	Sy	4	11	8	3	1	13	3	3	25	7,3	348	77
Bräkne-Hoby	BH	46	75	70	18	33	32	7	6	25	4,4	0	40
Hammenhög	Ha	48	100	96	23	51	53	12	8	31	4,4	0	45
Alnarp	Al	53	118	92	28	45	77	20	9	93	4,7	0	42
Hilleshög	Hi	32	78	68	20	41	36	12	6	27	4,4	0	54
Ås	Ås	5	13	13	3	1	35	4	3	29	6,6	185	67
Vågåmo	Vå	2	—	—	—	—	—	—	—	—	6,9	120	32
Lista	Li	6	27	390	3	2	242	14	34	24	6,5	72	282
Kauhava	Ka	4	8	6	1	1	6	5	1	12	6,4	59	38
Kuopio	Ku	19	12	5	1	1	8	5	1	12	6,1	21	10
Jyväskylä	Jy	8	8	5	1	1	5	3	1	12	6,3	91	16
Tvärminne	Tv	8	18	8	3	3	9	4	1	15	6,1	37	31
Ödum	Öd	7	13	15	3	4	10	2	1	6	4,8	0	36
Askov	As	39	101	280	22	47	169	15	16	40	4,4	0	75
Tystofte	Ty	36	79	101	17	49	81	12	11	48	5,9	20	41

L 501

Station	Code	$\mu g/m^3$ (= kg/km^3)						
		S	Cl	NH_3-N	Na	K	Mg	Ca
Riksgränsen	Ri	—	—	—	—	—	—	—
Kiruna	Ki	4.6	4.9	2.1	3.9	7.6	0.0	6.4
Arjeplog	Ar	—	—	—	—	—	—	—
Öjebyn	Öj	—	—	—	—	—	—	—
Röbäcksdalen	Rö	5.3	0.0	1.1	1.1	0.2	0.0	0.9
Offer	Of	1.8	0.3	2.2	2.1	0.7	0.0	2.2
Bredkälen	Br	1.2	4.0	1.6	2.7	0.0	3.2	1.6
Åre	ÅF	—	—	—	—	—	—	—
Åre, Hummeln	ÅH	—	—	—	—	—	—	—
Sveg	Sv	—	—	—	—	—	—	—
Mora	Mo	14.5	2.1	3.8	0.8	0.7	0.0	2.6
Åmot	Åm	4.6	3.7	1.6	2.2	0.7	0.0	4.2
Sala	Sa	Sample spoiled (dry)						
Ultuna	Ul	9.9	2.3	1.3	0.4	0.0	0.1	0.5
Erken	Er	7.0	2.6	1.0	1.3	0.1	1.6	1.5
Strängnäs	St	10.3	3.2	2.5	0.7	0.2	0.0	1.0
Forshult	Fo	—	—	—	—	—	—	—
Kvarntorp	Kv1	107.9	9.7	11.7	4.1	5.8	1.5	19.3
Kvarntorp	Kv7	4.1	0.3	2.0	0.4	0.0	0.1	0.9
Västra Ny	VN	—	—	—	—	—	—	—
Lanna	La	8.1	8.0	3.1	3.1	0.3	0.0	3.2
Bornö	Bo	6.9	29.3	1.9	11.0	0.2	1.6	5.8
Vinga	Vi	8.7	78.4	1.6	42.1	1.5	4.5	3.1
Falsterbobruk	Fa	4.2	5.2	1.1	1.1	0.1	0.0	2.4
Flahult	Fl	2.1	6.1	1.2	1.2	0.0	2.2	2.3
Ambjörnarp	Am	—	—	—	—	—	—	—
Simlångsdalen	Si	—	—	—	—	—	—	—
Plönninge	Pl	11.1	4.5	4.7	1.3	0.0	0.7	1.8
Söraby	Sö	—	—	—	—	—	—	—
Smedby	Sm	5.4	0.9	1.6	0.7	0.2	0.5	1.3
Sylfaste	Sy	4.3	5.2	1.7	1.8	0.2	0.4	1.2
Bräkne-Hoby	BH	—	—	—	—	—	—	—
Hammenhög	Ha	—	—	—	—	—	—	—
Alnarp	Al	20.2	2.5	2.8	1.1	0.1	0.7	2.8
Hilleshög	Hi	—	—	—	—	—	—	—
Ås	Ås	10.5	0.0	2.4	1.3	0.0	0.6	1.5
Vågåmo	Vå	Sample spoiled (dry)						
Lista	Li	6.8	66.3	1.1	33.9	1.7	6.5	2.4
Kauhava	Ka	6.6	4.8	2.6	3.2	2.4	1.7	1.9
Kuopio	Ku	—	—	—	—	—	—	—
Jyväskylä	Jy	1.6	1.7	1.1	1.7	0.5	1.2	2.1
Tvärminne	Tv	5.2	4.9	1.3	2.3	1.2	0.6	2.2
Ödum	Öd	24.7	15.8	8.1	3.4	0.0	0.7	5.3
Askov	As	—	—	—	—	—	—	—
Tystofte	Ty	16.2	15.6	5.0	7.7	0.5	3.3	4.1

L 502

Station	Code	$\mu g/m^3$ (= kg/km^3)						
		S	Cl	NH_3-N	Na	K	Mg	Ca
Riksgränsen	Ri	—	—	—	—	—	—	—
Kiruna	Ki	—	—	—	—	—	—	—
Arjeplog	Ar	—	—	—	—	—	—	—
Öjebyn	Öj	—	—	Sample spoiled (dry)			—	—
Röbäcksdalen	Rö	9.0	0.3	1.3	0.9	0.5	0.2	1.6
Offer	Of	0.8	4.0	2.0	0.7	1.0	0.4	3.4
Bredkälen	Br	0.0	3.8	1.2	1.5	0.6	0.9	1.9
Åre	ÅF	5.3	4.3	0.9	1.7	1.2	0.1	1.3
Åre, Hummeln	ÅH	5.0	5.4	1.2	2.8	2.3	0.8	5.7
Sveg	Sv	—	—	—	—	—	—	—
Mora	Mo	13.9	3.3	2.6	1.1	0.7	0.0	2.0
Åmot	Åm	4.3	2.7	1.5	1.0	0.4	0.9	2.3
Sala	Sa	11.1	2.8	2.6	3.8	1.7	0.5	6.5
Ultuna	Ul	11.0	3.8	1.3	0.6	0.5	1.4	0.7
Erken	Er	7.4	0.0	1.8	0.9	0.8	3.6	1.3
Strängnäs	St	7.6	1.5	2.4	0.7	0.8	0.3	2.2
Forshult	Fo	3.4	0.0	1.2	1.2	2.2	3.0	4.0
Kvarntorp	Kv1	89.3	6.7	10.5	4.1	8.3	4.0	51.1
Kvarntorp	Kv7	4.3	1.1	1.5	0.4	0.5	0.2	2.2
Västra Ny	VN	—	—	—	—	—	—	—
Lanna	La	—	—	—	—	—	—	—
Bornö	Bo	—	—	—	—	—	—	—
Vinga	Vi	—	—	—	—	—	—	—
Falsterbobruk	Fa	11.1	10.7	3.2	1.0	0.5	1.1	2.2
Flahult	Fl	3.1	1.7	1.1	0.6	0.2	0.2	3.4
Ambjörnarp	Am	—	—	—	—	—	—	—
Simlångsdalen	Si	—	—	—	—	—	—	—
Plönninge	Pl	9.6	3.3	15.5	1.0	0.0	1.6	4.9
Söraby	Sö	—	—	—	—	—	—	—
Smedby	Sm	4.4	0.0	1.6	0.5	0.0	0.0	2.3
Sylfaste	Sy	—	—	Sample spoiled (dry)			—	—
Bräkne-Hoby	BH	—	—	—	—	—	—	—
Hammenhög	Ha	—	—	—	—	—	—	—
Alnarp	Al	19.3	1.0	2.8	0.9	0.1	0.6	2.7
Hilleshög	Hi	—	—	—	—	—	—	—
Ås	Ås	4.4	0.0	0.7	0.4	0.1	0.1	1.8
Vågåmo	Vå	0.3	0.0	5.8	0.8	0.5	0.3	4.4
Lista	Li	3.9	31.4	1.1	11.3	1.0	2.1	2.7
Kauhava	Ka	2.8	0.0	1.8	1.8	1.9	1.0	8.8
Kuopio	Ku	1.7	3.7	1.0	0.4	0.2	2.2	7.0
Jyväskylä	Jy	6.7	2.4	1.7	0.3	0.2	5.6	5.2
Tvärminne	Tv	4.9	0.0	1.4	0.7	0.1	4.5	3.8
Ödum	Öd	15.1	0.0	3.2	1.8	0.0	0.8	8.3
Askov	As	—	—	—	—	—	—	—
Tystofte	Ty	20.6	1.8	5.9	1.5	0.2	0.3	4.7

L 503

Station	Code	$\mu g/m^3$ (= kg/km^3)						
		S	Cl	NH_3-N	Na	K	Mg	Ca
Riksgränsen	Ri	16.5	16.9	6.2	11.1	3.1	0.7	11.9
Kiruna	Ki	3.9	2.9	0.9	0.8	9.0	0.2	3.8
Arjeplog	Ar	—	—	—	—	—	—	—
Öjebyn	Öj	—	—	Sample spoiled (dry)			—	—
Röbäcksdalen	Rö	11.0	0.6	1.7	0.6	0.2	0.2	2.0
Offer	Of	6.4	0.0	2.5	0.7	0.0	0.4	2.8
Bredkälen	Br	4.7	0.9	0.9	1.6	0.7	0.5	3.3
Åre	ÅF	4.7	7.0	1.0	0.8	0.9	0.2	2.3
Åre, Hummeln	ÅH	2.1	5.5	0.7	1.6	0.3	0.5	3.4
Sveg	Sv	—	—	—	—	—	—	—
Mora	Mo	7.1	2.1	2.7	0.4	0.2	0.1	1.2
Åmot	Åm	—	—	—	—	—	—	—
Sala	Sa	—	—	Sample spoiled (dry)			—	—
Ultuna	Ul	8.7	2.0	1.6	0.4	0.1	2.9	0.6
Erken	Er	6.0	0.0	1.2	0.6	0.1	0.3	1.1
Strängnäs	St	9.0	2.5	2.9	1.2	0.0	0.6	1.3
Forshult	Fo	6.5	0.0	1.5	1.3	0.5	0.3	4.2
Kvarntorp	Kv1	71.9	7.4	10.4	2.0	10.8	3.8	32.0
Kvarntorp	Kv7	7.9	0.2	2.8	1.7	0.3	0.3	2.0
Västra Ny	VN	—	—	—	—	—	—	—
Lanna	La	—	—	—	—	—	—	—
Bornö	Bo	4.7	20.0	10.0	1.2	0.8	0.3	3.2
Vinga	Vi	16.5	178	2.4	79.6	6.6	13.8	11.4
Falsterbobruk	Fa	6.7	0.0	1.1	2.3	2.4	1.3	9.0
Flahult	Fl	6.9	5.9	8.2	2.2	2.6	1.1	2.8
Ambjörnarp	Am	—	—	—	—	—	—	—
Simlångsdalen	Si	—	—	—	—	—	—	—
Plönninge	Pl	8.8	4.7	2.0	1.0	0.0	0.4	2.4
Söraby	Sö	—	—	—	—	—	—	—
Smedby	Sm	5.2	0.8	1.4	0.6	0.0	0.0	0.7
Sylfaste	Sy	—	—	Sample spoiled (dry)			—	—
Bräkne-Hoby	BH	—	—	—	—	—	—	—
Hammenhög	Ha	—	—	—	—	—	—	—
Alnarp	Al	21.0	3.6	3.3	1.7	0.4	0.4	3.0
Hilleshög	Hi	—	—	—	—	—	—	—
Ås	Ås	12.4	2.8	2.4	1.8	0.0	0.8	4.0
Vågåmo	Vå	4.2	5.8	7.1	2.0	0.3	0.5	5.1
Lista	Li	6.5	62.5	1.6	24.8	0.8	4.0	4.7
Kauhava	Ka	6.5	0.0	2.8	0.2	0.0	2.7	4.0
Kuopio	Ku	4.9	1.4	1.3	0.0	0.0	0.4	4.0
Jyväskulä	Jy	8.0	13.4	4.0	2.7	1.5	1.8	8.1
Tvärminne	Tv	6.4	0.2	1.5	0.3	0.0	0.8	4.4
Ödum	Öd	7.4	1.7	3.0	0.7	0.0	0.5	1.1
Askov	As	—	—	—	—	—	—	—
Tystofte	Ty	13.7	8.5	6.7	1.9	0.4	0.7	4.0

Uppsala and Stockholm in April 1955.

H. Egner E. Eriksson
Agricultural College, Ultuna and University of Stockholm

Routine Forecasting with the Barotropic Model

by

P. BERGTHORSSON, Vedurstofan, Reykjavik, B. R. DÖÖS, University of Stockholm, S. FRYKLUND, The Swedish Weather Service, Stockholm, O. HAUG, Det Norske Meteorologiske Institutt, Oslo, and R. LINDQUIST, The Royal Swedish Air Force.

(Manuscript received May 25, 1955)

Introduction

Since 1 December 1954 the Royal Swedish Air Force Weather Service has been conducting a series of test forecasts of the 500-mb topography under operational conditions. Forecasts for 24, 48 and 72 hours have been prepared on a routine basis, using the barotropic model and the digital computer BESK. The tests to date have covered the periods 1—16 December 1954, 17 January—25 February 1955 and from 12 April 1955 to the time of writing. It was originally planned to issue 42 sets of three forecasts during the first two periods, but some forecasts failed due to minor breakdowns of the computer (see tables).

The daily schedule for preparing the numerical forecasts has normally proceeded according to the following plan: The analysis of the 0300 GMT 500-mb map is completed at about 1230 GMT (the approximate time of receipt of the last observations). The reading of the data from the grid points and the punching and checking of the input data tape for BESK follows, and is completed about 1530 GMT. The actual computation follows immediately and is normally completed after 65 to 75 minutes.

Methods of verification

The verification of the prognostic charts have been made for the area shown in fig. 1.

Two different methods of verification were applied. In the first place the correlation coefficients and the errors using the observed and computed changes in all grid points within the verification area. Secondly, in order to get a direct estimate of the prognostic value of the charts to the forecaster, a subjective verification was carried out by 6 meteorologists from 4 different weather services (Vedurstofan, Iceland, Det Norske Meteorologiske Institutt, Norway, the Swedish Weather Service (SMHI) and the Royal Swedish Air Force Weather Service). The prognostic value of the charts within the verification area was given in the form of numbers from 1 to 5, number 1 denoting "failing", and 5 "very good". The natural way of getting an estimate of the value of numerical weather prediction in routine forecasting is to compare the results with the corresponding results obtained by conventional methods. In order to obtain a comparison of this kind, the two methods of verification mentioned above were also applied to the corresponding series of 24 and 48 hrs prognostic charts issued by the Swedish Weather Service (SMHI). The results of the forecasts are given in table I, II and III and the averages are also presented in fig. 2.

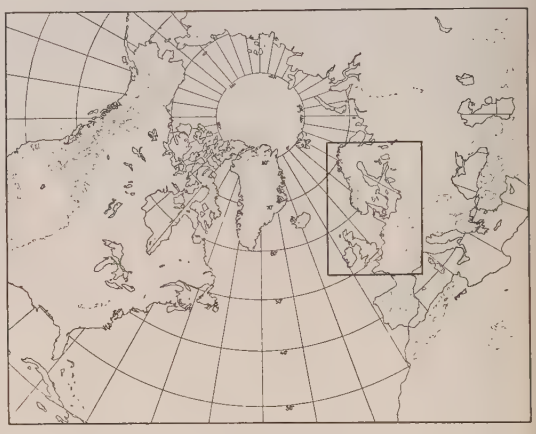


Fig. 1. The forecast area and the verification area.

Table I

24 hrs forecasts							
Date of initial map		Numerical method			Conventional method		
		s	r	ε	s	r	ε
1 Dec.	-54, 03..	5	0.92	50	4	0.61	86
2 »	-54, 03..	5	0.89	102	5	0.90	47
3 »	-54, 03..	4	0.73	93	4	0.66	108
6 »	-54, 03..	5	0.93	18	3	0.63	65
7 »	-54, 03..	3	0.75	68	3	0.79	58
8 »	-54, 03..	5	0.94	49	4	0.71	81
9 »	-54, 03..	5	0.79	69	3	0.00	288
10 »	-54, 03..	5	0.90	62	5	0.83	48
13 »	-54, 03..	5	0.75	51	4	0.06	102
15 »	-54, 03..	4	0.76	71	4	0.83	86
16 »	-54, 03..	5	0.85	56	5	0.85	65
17 »	-54, 03..	5	0.92	60	5	0.91	48
17 Jan.	-55, 03..	4	0.88	52	3	0.76	67
18 »	-55, 03..	4	0.89	46	1	0.51	74
19 »	-55, 03..	3	0.36	74	3	0.84	68
24 »	-55, 03..	3	0.70	51	3	0.56	65
27 »	-55, 03..	5	0.74	38	5	0.90	36
28 »	-55, 03..	5	0.66	41	5	0.36	59
31 »	-55, 03..	5	0.91	24	5	0.69	49
1 Febr.	-55, 03..	5	0.81	47	4	0.66	60
2 »	-55, 03..	3	0.53	53	4	0.78	31
3 »	-55, 03..	5	0.91	26	4	0.68	43
4 »	-55, 03..	4	0.75	34	4	0.43	55
7 »	-55, 03..	4	0.82	46	4	0.65	54
8 »	-55, 03..	3	0.76	39	2	0.31	76
9 »	-55, 03..	5	0.77	48	4	0.80	59
10 »	-55, 03..	4	0.76	51	4	0.63	40
11 »	-55, 03..	3	0.71	78	5	0.81	49
14 »	-55, 03..	2	0.65	79	4	0.85	47
15 »	-55, 03..	2	0.80	55	2	0.72	83
16 »	-55, 03..	4	0.76	81	3	0.39	114
17 »	-55, 03..	4	0.90	78	4	0.91	62
18 »	-55, 03..	5	0.23	118	5	0.65	47
22 »	-55, 03..	4	0.74	59	2	0.20	64
23 »	-55, 03..	4	0.78	40	2	0.35	68
24 »	-55, 03..	5	0.72	51	5	0.22	70
25 »	-55, 03..	5	0.83	39	3	0.92	50
	Mean	4.2	0.77	57	3.8	0.63	70

Table II

48 hrs forecasts							
Date of initial map		Numerical method			Conventional method		
		s	r	ε	s	r	ε
1 Dec.	-54, 03..	3	0.79	134	5	0.90	113
2 »	-54, 03..	3	0.75	174	3	0.52	109
3 »	-54, 03..	2	0,71	237	1	0.48	187
6 »	-54, 03..	3	0.71	118	4	0.53	157
7 »	-54, 03..	3	0.77	114	2	0.66	110
8 »	-54, 03..	4	0.92	112	3	— 0.24	309
9 »	-54, 03..	3	0.63	200	2	0.59	178
10 »	-54, 03..	4	0.41	245	4	0.78	77
13 »	-54, 03..	2	0.60	142	2	0.53	107
15 »	-54, 03..	3	0.02	169	3	0.02	150
16 »	-54, 03..	4	0.86	100	4	0.61	87
17 »	-54, 03..	5	0.88	113	3	0.66	130
17 Jan.	-55, 03..	3	0.61	99	3	0.62	82
18 »	-55, 03..	—	—	—	—	—	—
19 »	-55, 03..	3	0.65	81	4	0.80	145
24 »	-55, 03..	—	—	—	—	—	—
27 »	-55, 03..	5	0,13	96	4	0.48	65
28 »	-55, 03..	5	0.21	97	5	0.66	53
31 »	-55, 03..	5	0.93	53	3	0.68	69
1 Febr.	-55, 03..	3	0.72	100	4	0.75	60
2 »	-55, 03..	4	0.78	72	4	0.77	63
3 »	-55, 03..	4	0.87	61	1	0.65	23
4 »	-55, 03..	5	0.80	63	2	0.22	91
7 »	-55, 03..	4	0.87	55	4	0.64	91
8 »	-55, 03..	3	0.77	62	4	0.83	100
9 »	-55, 03..	4	0.77	77	2	0.48	114
10 »	-55, 03..	4	0.86	134	2	0.61	64
11 »	-55, 03..	2	0.38	191	3	0.45	84
14 »	-55, 03..	3	0.43	188	4	0.32	127
15 »	-55, 03..	5	0.49	148	4	0.48	143
16 »	-55, 03..	—	—	—	—	—	—
17 »	-55, 03..	5	0.93	136	3	0.76	114
18 »	-55, 03..	4	0.39	119	4	0.53	85
22 »	-55, 03..	3	0.78	69	2	0.34	93
23 »	-55, 03..	3	0.71	67	3	0.02	114
24 »	-55, 03..	5	0.63	104	1	0.14	105
25 »	-55, 03..	4	0.66	89	1	0.77	73
	Mean	3.7	0.66	118	3.0	0.53	108

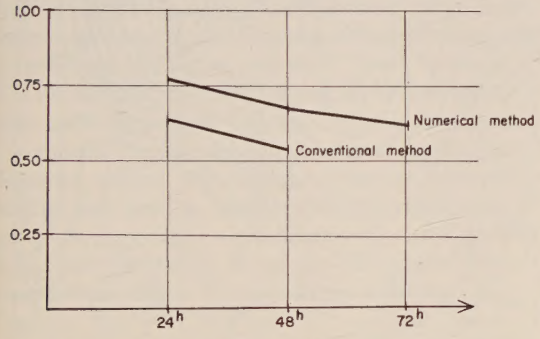


Fig. 2. Graphical representation of the variation of the correlation coefficient.

Practical application in daily weather forecasting

The numerical prognostic charts have been used on a daily routine basis for forecasting the general weather situation for the next 3—5 days. In this connection, the 72-hr forecasts have been of special interest, as prognostic charts for as much as 3 days have not previously been used at the Swedish Air Force Weather Service.

In spite of the systematic errors which occasionally appear in the 72-hr forecasts (Bolin, 1955) they have in most cases been very useful for the routine weather forecasting.

Table III

72 hrs forecasts			
Date of initial map	Numerical method		
	s	r	ε
1 Dec. -54, 03	3	0.83	192
2 » -54, 03	3	0.67	163
3 » -54, 03	—	0.77	253
6 » -54, 03	—	—	—
7 » -54, 03	4	0.76	209
8 » -54, 03	3	0.68	180
9 » -54, 03	—	—	—
10 » -54, 03	3	— 0.09	345
13 » -54, 03	4	0.85	191
15 » -54, 03	3	0.71	182
16 » -54, 03	3	0.61	153
17 » -54, 03	3	0.90	224
17 Jan. -55, 03	2	0.31	153
18 » -55, 03	—	—	—
19 » -55, 03	—	—	—
24 » -55, 03	—	—	—
27 » -55, 03	4	0.44	152
28 » -55, 03	3	— 0.23	188
31 » -55, 03	5	0.90	84
1 Febr. -55, 03	2	0.73	128
2 » -55, 03	4	0.74	151
3 » -55, 03	4	0.76	118
4 » -55, 03	3	0.76	105
7 » -55, 03	2	0.68	98
8 » -55, 03	2	0.77	109
9 » -55, 03	4	0.89	165
10 » -55, 03	1	0.42	234
11 » -55, 03	—	—	—
14 » -55, 03	—	—	—
15 » -55, 03	3	0.26	281
16 » -55, 03	—	—	—
17 » -55, 03	3	0.68	160
18 » -55, 03	—	—	—
22 » -55, 03	1	0.53	107
23 » -55, 03	—	—	—
24 » -55, 03	3	0.60	196
25 » -55, 03	3	0.56	222
Mean	3.0	0.61	176

An example of a successful 72-hr forecast is shown in fig. 2. The weather situation for several days prior to the date when this forecast was made had been characterised by a N to NW flow and cold weather over Scandinavia. During the 3 following days the weather situation changed, resulting in a milder weather-type dominated by a SW flow over southern Scandinavia. This development was very well indicated by the 72-hr forecast.

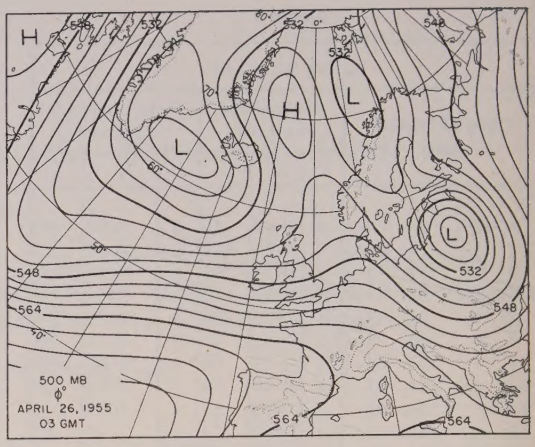


Fig. 3 a. Observed 500 mb contours on April 26, 1955, 0300 GMT. The heights are given in decameter as unit.

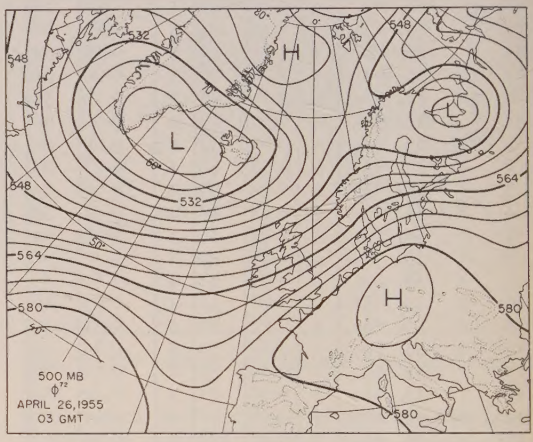


Fig. 3 b. 72-hour forecast of 500 mb from map shown in a.

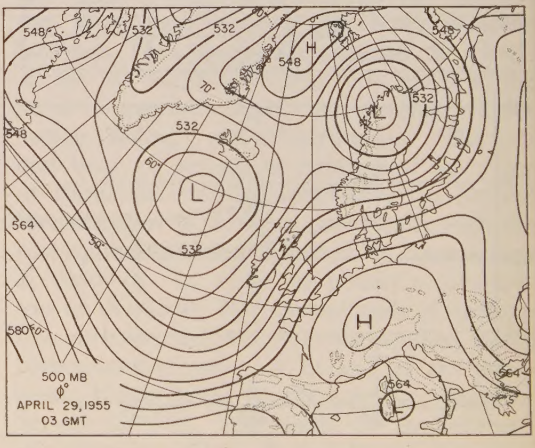


Fig. 3 c. Observed contours on April 29, 1955, 0300 GMT.

REFERENCE

BOLIN, B., 1955: Numerical forecasting with the barotropic model. *Tellus* **7**, 27—49.

On «Results of forecasting with the barotropic model»

Dear Sir,

In a recent paper describing results of forecasting by numerical methods (Staff Members, University of Stockholm, 1954), it was stated that the correlation between observed and computed changes is not a satisfactory measure of the goodness of a forecast. Previously the same group (Staff Members, University of Stockholm, 1952) had raised the objection that a correlation coefficient measures the similarity in behaviour of two variables without any reference to systematic errors such as a difference in scale of the variations. However there is a more fundamental objection which in no way depends on the properties of a correlation coefficient as such, but applies to any comparison of observed and computed *changes*, whether quantitative or qualitative.

The basis of this objection is illustrated by a hypothetical theory which yields "computed" changes in every case equal and opposite to the corresponding initial departure of the forecast element from its normal value. Any examination of such "computed" changes and those observed would on the whole show considerable agreement between them – it can be shown theoretically that without persistence one would expect a correlation of about 0.7 between observed and "computed" changes, while somewhat smaller values would appear if persistence were important. The "forecasts" resulting from this "theory" are of course identical with the ordinary climatological forecasts, which can be made without any theoretical background. They therefore do not represent any practical progress in spite of the good agreement between observed and "computed" changes.

Now it may be argued that the main purpose of many experimental forecasts is to obtain confirmation or otherwise of a particular theory, and that agreement between observed and computed changes is sufficient from this point of view. However, for the forecast described above, it can be shown that good agreement will also result if the same "computed" changes are compared with the changes observed during any arbitrary forecast interval. In this case there is good agreement without any apparent theoretical justification, so that the good agreement previously found for the proper forecast interval cannot be regarded as confirming the "theory".

Reflection shows that the truth lies between these two positions of extreme optimism and extreme pessimism in the interpretation of the results. Thus if a real physical theory existed which could predict on every occasion a return to normal conditions (which is the case for the hypothetical theory considered), the very important problem of explaining these normal or mean conditions (in effect the problem of the general circulation) would have been solved. However such a theory obviously plays no part in explaining *variations* from mean conditions, which might be considered as the main task of any forecast theory.

The hypothetical case therefore shows good agreement between observed and computed *changes* associated with a "theory" which is of no practical value and which is satisfactory only in a very restricted sense. While the example given is admittedly rather artificial, similar reasoning applied to any other forecast methods will show that that part of the agreement between observation and theory which reflects a "return to normal" effect in the underlying theory will be similarly without practical value. It may be valuable as a contribution to the general circulation problem, but not to the genuine forecast problem of explaining variations from normal conditions.

It follows from this that the relationship between observed and computed *changes* (whether expressed as a correlation or in any other way) cannot be interpreted directly in terms of goodness of a forecast. Even in comparing forecasts by different methods or under different conditions, the comparison usually made (i.e. by the magnitude of the correlation) would have meaning only if the "return to normal" effect contributed equally to all types of forecast. However the possibility cannot be ruled out that an apparently better method of forecasting contains in reality only a stronger tendency for "return to normal", so that the improvement is largely illusory.

This pitfall can be avoided by using a partial correlation coefficient in which effects of variation in the initial departure from normal have been eliminated. The partial correlation, given by $r_p = \frac{r - r_1 r_2}{\{(1 - r_1^2)(1 - r_2^2)\}^{\frac{1}{2}}}$ where r is the correlation between observed and computed changes and r_1 and r_2 are correlations between each of these quantities and the initial departure from normal, depends only on that part of the relationship between observed and computed changes which is inde-

pendent of the "return to normal" effect. It is therefore a much more satisfactory measure of the goodness of a forecast than is the simple correlation r between observed and computed changes.

The importance of the suggested change inverification procedure when applied to numerical forecasts can be judged only by calculating r_p from the original data, a course unfortunately not available to the present writer. However r_p may differ considerably from the simple correlation r, so that values of the latter in published verifications give little information regarding the true contributions of the various theoretical models towards solution of the forecast problem.

REFERENCES

STAFF MEMBERS, *The Institute of Meteorology, University of Stockholm*, 1952: Preliminary report on the prognostic value of barotropic models in the forecasting of 500 mb height changes. *Tellus*, **4**, 21—30.

STAFF MEMBERS, *The Institute of Meteorology, University of Stockholm*, 1954: Results of forecasting with the barotropic model on an electronic computer (BESK). *Tellus* **6**, 139—149.

January 27, 1955

ALISON M. GRANT
University of Melbourne, Australia

Miss Grant's comments on the verification of forecasts, conventional or numerical, are interesting. The following additional remarks should be given in connection to the numerical forecasts made at the Institute of Meteorology, University of Stockholm.

The partial correlation coefficients have been computed for the 24, 48 and 72 hour forecasts from October 2, 1954, 0300 GCT (cf. BOLIN, 1955). They were 0.89 (0.92), 0.84 (0.92) and 0.79 (0.87), where the figures within the parentheses are the corresponding ordinary correlation coefficients. It is seen that the partial correlation coefficients are consistently lower as would be expected. The coefficients r_1 and r_2 varied between 0.55 and 0.70 in this case. These values of r_1 and r_2 give $r_p = 0$, if $r \approx 0.4$.

For a detailed comparison between various forecasts one probably should also include the partial correlation coefficient. Under the assumption, however, that the persistence of atmospheric flow pattern does not vary too much (which of course is somewhat doubtful), it merely means an expansion of the scale from $r = 0.4$—0.5 meaning a worthless forecast to $r = 1$ meaning a perfect forecast to a scale $r_p = 0$ = failure to $r_p = 1$ = perfect.

It should finally be stressed that any correlation coefficient of this kind gives a very incomplete description of the success or failure of a forecast. This is most clearly seen from the experiment of subjective verification reported in a note in this issue of Tellus (pp. 272—274). Quantities as error, standard deviations, etc. may be of some additional help, but a forecast map is mainly used for evaluating changes of the *flow*, and one should therefore aim at a method which gives some information about the success in this respect.

REFERENCE

BOLIN, B., 1955: Numerical forecasting with the barotropic model. *Tellus*, **7**, 27—49.

BERT BOLIN
Institute of Meteorology
University of Stockholm